18.600: Lecture 17

Continuous random variables

Scott Sheffield

MIT

Outline

Continuous random variables

Expectation and variance of continuous random variables

Uniform random variable on $[0, 1]$

Uniform random variable on $[\alpha, \beta]$

Measurable sets and a famous paradox

Outline

Continuous random variables

Expectation and variance of continuous random variables

Uniform random variable on $[0, 1]$

Uniform random variable on $[\alpha, \beta]$

Measurable sets and a famous paradox

Continuous random variables

- Say X is a **continuous random variable** if there exists a **probability density function** $f = f_X$ on $\mathbb{R}$ such that $P\{X \in B\} = \int_B f(x)dx := \int 1_B(x)f(x)dx$.

Continuous random variables

- Say X is a **continuous random variable** if there exists a **probability density function** $f = f_X$ on $\mathbb{R}$ such that $P\{X \in B\} = \int_B f(x)dx := \int 1_B(x)f(x)dx$.
- We may assume $\int_{\mathbb{R}} f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1$ and f is non-negative.

Continuous random variables

- Say X is a **continuous random variable** if there exists a **probability density function** $f = f_X$ on $\mathbb{R}$ such that $P\{X \in B\} = \int_B f(x)dx := \int 1_B(x)f(x)dx$.
- We may assume $\int_{\mathbb{R}} f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1$ and f is non-negative.
- Probability of interval $[a, b]$ is given by $\int_a^b f(x)dx$, the area under f between a and b.

Continuous random variables

- Say X is a **continuous random variable** if there exists a **probability density function** $f = f_X$ on $\mathbb{R}$ such that $P\{X \in B\} = \int_B f(x)dx := \int 1_B(x)f(x)dx$.
- We may assume $\int_{\mathbb{R}} f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1$ and f is non-negative.
- Probability of interval $[a, b]$ is given by $\int_a^b f(x)dx$, the area under f between a and b.
- Probability of any single point is zero.

Continuous random variables

- Say X is a **continuous random variable** if there exists a **probability density function** $f = f_X$ on $\mathbb{R}$ such that $P\{X \in B\} = \int_B f(x)dx := \int 1_B(x)f(x)dx$.
- We may assume $\int_{\mathbb{R}} f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1$ and f is non-negative.
- Probability of interval $[a, b]$ is given by $\int_a^b f(x)dx$, the area under f between a and b.
- Probability of any single point is zero.
- Define **cumulative distribution function** $F(a) = F_X(a) := P\{X < a\} = P\{X \leq a\} = \int_{-\infty}^{a} f(x)dx$.

Simple example

- Suppose $f(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$

Simple example

- Suppose $f(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?

Simple example

- Suppose $f(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?
- What is $P\{X = 3/2\}$?

Simple example

- Suppose $f(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?
- What is $P\{X = 3/2\}$?
- What is $P\{1/2 < X < 3/2\}$?

Simple example

- Suppose $f(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?
- What is $P\{X = 3/2\}$?
- What is $P\{1/2 < X < 3/2\}$?
- What is $P\{X \in (0,1) \cup (3/2,5)\}$?

Simple example

- Suppose $f(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?
- What is $P\{X = 3/2\}$?
- What is $P\{1/2 < X < 3/2\}$?
- What is $P\{X \in (0,1) \cup (3/2,5)\}$?
- What is F?

Simple example

- Suppose $f(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?
- What is $P\{X = 3/2\}$?
- What is $P\{1/2 < X < 3/2\}$?
- What is $P\{X \in (0,1) \cup (3/2,5)\}$?
- What is F?
- $F(a) = F_X(a) = \begin{cases} 0 & a \leq 0 \\ a/2 & 0 < a < 2 \\ 1 & a \geq 2 \end{cases}$

Simple example

- Suppose $f(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?
- What is $P\{X = 3/2\}$?
- What is $P\{1/2 < X < 3/2\}$?
- What is $P\{X \in (0,1) \cup (3/2,5)\}$?
- What is F?
- $F(a) = F_X(a) = \begin{cases} 0 & a \le 0 \\ a/2 & 0 < a < 2 \\ 1 & a \ge 2 \end{cases}$
- In general $P(a \le x \le b) = F(b) - F(x)$.

Simple example

- Suppose $f(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?
- What is $P\{X = 3/2\}$?
- What is $P\{1/2 < X < 3/2\}$?
- What is $P\{X \in (0,1) \cup (3/2,5)\}$?
- What is F?
- $F(a) = F_X(a) = \begin{cases} 0 & a \leq 0 \\ a/2 & 0 < a < 2 \\ 1 & a \geq 2 \end{cases}$
- In general $P(a \leq x \leq b) = F(b) - F(x)$.
- We say that X is **uniformly distributed on** $[0,2]$.

Another example

- Suppose $f(x) = \begin{cases} x/2 & x \in [0,2] \\ 0 & 0 \notin [0,2]. \end{cases}$

Another example

- Suppose $f(x) = \begin{cases} x/2 & x \in [0,2] \\ 0 & 0 \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?

Another example

- Suppose $f(x) = \begin{cases} x/2 & x \in [0,2] \\ 0 & 0 \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?
- What is $P\{X = 3/2\}$?

Another example

- Suppose $f(x) = \begin{cases} x/2 & x \in [0,2] \\ 0 & 0 \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?
- What is $P\{X = 3/2\}$?
- What is $P\{1/2 < X < 3/2\}$?

Another example

- Suppose $f(x) = \begin{cases} x/2 & x \in [0,2] \\ 0 & 0 \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?
- What is $P\{X = 3/2\}$?
- What is $P\{1/2 < X < 3/2\}$?
- What is F?

Another example

- Suppose $f(x) = \begin{cases} x/2 & x \in [0,2] \\ 0 & 0 \notin [0,2]. \end{cases}$
- What is $P\{X < 3/2\}$?
- What is $P\{X = 3/2\}$?
- What is $P\{1/2 < X < 3/2\}$?
- What is F?
- $F_X(a) = \begin{cases} 0 & a \leq 0 \\ a^2/4 & 0 < a < 2 \\ 1 & a \geq 2 \end{cases}$

Outline

Continuous random variables

Expectation and variance of continuous random variables

Uniform random variable on $[0,1]$

Uniform random variable on $[\alpha, \beta]$

Measurable sets and a famous paradox

Outline

Expectations of continuous random variables

- Recall that when X was a discrete random variable, with $p(x) = P\{X = x\}$, we wrote

$$E[X] = \sum_{x:p(x)>0} p(x)x.$$

Expectations of continuous random variables

- Recall that when X was a discrete random variable, with $p(x) = P\{X = x\}$, we wrote

$$E[X] = \sum_{x:p(x)>0} p(x)x.$$

- How should we define $E[X]$ when X is a continuous random variable?

Expectations of continuous random variables

- Recall that when X was a discrete random variable, with $p(x) = P\{X = x\}$, we wrote

$$E[X] = \sum_{x:p(x)>0} p(x)x.$$

- How should we define $E[X]$ when X is a continuous random variable?
- Answer: $E[X] = \int_{-\infty}^{\infty} f(x)xdx$.

Expectations of continuous random variables

- Recall that when X was a discrete random variable, with $p(x) = P\{X = x\}$, we wrote

$$E[X] = \sum_{x:p(x)>0} p(x)x.$$

- How should we define $E[X]$ when X is a continuous random variable?
- Answer: $E[X] = \int_{-\infty}^{\infty} f(x)xdx$.
- Recall that when X was a discrete random variable, with $p(x) = P\{X = x\}$, we wrote

$$E[g(X)] = \sum_{x:p(x)>0} p(x)g(x).$$

Expectations of continuous random variables

- Recall that when X was a discrete random variable, with $p(x) = P\{X = x\}$, we wrote

$$E[X] = \sum_{x:p(x)>0} p(x)x.$$

- How should we define $E[X]$ when X is a continuous random variable?
- Answer: $E[X] = \int_{-\infty}^{\infty} f(x)xdx$.
- Recall that when X was a discrete random variable, with $p(x) = P\{X = x\}$, we wrote

$$E[g(X)] = \sum_{x:p(x)>0} p(x)g(x).$$

- What is the analog when X is a continuous random variable?

Expectations of continuous random variables

- Recall that when X was a discrete random variable, with $p(x) = P\{X = x\}$, we wrote

$$E[X] = \sum_{x:p(x)>0} p(x)x.$$

- How should we define $E[X]$ when X is a continuous random variable?
- Answer: $E[X] = \int_{-\infty}^{\infty} f(x)xdx$.
- Recall that when X was a discrete random variable, with $p(x) = P\{X = x\}$, we wrote

$$E[g(X)] = \sum_{x:p(x)>0} p(x)g(x).$$

- What is the analog when X is a continuous random variable?
- Answer: we will write $E[g(X)] = \int_{-\infty}^{\infty} f(x)g(x)dx$.

Variance of continuous random variables

- Suppose X is a continuous random variable with mean μ.

Variance of continuous random variables

- Suppose X is a continuous random variable with mean μ.
- We can write $\mathrm{Var}[X] = E[(X - \mu)^2]$, same as in the discrete case.

Variance of continuous random variables

- Suppose X is a continuous random variable with mean μ.
- We can write $\mathrm{Var}[X] = E[(X-\mu)^2]$, same as in the discrete case.
- Next, if $g = g_1 + g_2$ then $E[g(X)] = \int g_1(x)f(x)dx + \int g_2(x)f(x)dx = \int \big(g_1(x) + g_2(x)\big) f(x)dx = E[g_1(X)] + E[g_2(X)]$.

Variance of continuous random variables

- Suppose X is a continuous random variable with mean μ.
- We can write $\mathrm{Var}[X] = E[(X-\mu)^2]$, same as in the discrete case.
- Next, if $g = g_1 + g_2$ then $E[g(X)] = \int g_1(x)f(x)dx + \int g_2(x)f(x)dx = \int \big(g_1(x) + g_2(x)\big) f(x)dx = E[g_1(X)] + E[g_2(X)]$.
- Furthermore, $E[ag(X)] = aE[g(X)]$ when a is a constant.

Variance of continuous random variables

- Suppose X is a continuous random variable with mean μ.
- We can write $\mathrm{Var}[X] = E[(X-\mu)^2]$, same as in the discrete case.
- Next, if $g = g_1 + g_2$ then $E[g(X)] = \int g_1(x)f(x)dx + \int g_2(x)f(x)dx = \int \big(g_1(x) + g_2(x)\big)f(x)dx = E[g_1(X)] + E[g_2(X)]$.
- Furthermore, $E[ag(X)] = aE[g(X)]$ when a is a constant.
- Just as in the discrete case, we can expand the variance expression as $\mathrm{Var}[X] = E[X^2 - 2\mu X + \mu^2]$ and use additivity of expectation to say that $\mathrm{Var}[X] = E[X^2] - 2\mu E[X] + E[\mu^2] = E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - E[X]^2$.

Variance of continuous random variables

- Suppose X is a continuous random variable with mean μ.
- We can write $\mathrm{Var}[X] = E[(X-\mu)^2]$, same as in the discrete case.
- Next, if $g = g_1 + g_2$ then $E[g(X)] = \int g_1(x)f(x)dx + \int g_2(x)f(x)dx = \int \big(g_1(x) + g_2(x)\big)f(x)dx = E[g_1(X)] + E[g_2(X)]$.
- Furthermore, $E[ag(X)] = aE[g(X)]$ when a is a constant.
- Just as in the discrete case, we can expand the variance expression as $\mathrm{Var}[X] = E[X^2 - 2\mu X + \mu^2]$ and use additivity of expectation to say that $\mathrm{Var}[X] = E[X^2] - 2\mu E[X] + E[\mu^2] = E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - E[X]^2$.
- This formula is often useful for calculations.

Outline

Continuous random variables

Expectation and variance of continuous random variables

Uniform random variable on $[0, 1]$

Uniform random variable on $[\alpha, \beta]$

Measurable sets and a famous paradox

Outline

Continuous random variables

Expectation and variance of continuous random variables

Uniform random variable on $[0, 1]$

Uniform random variable on $[\alpha, \beta]$

Measurable sets and a famous paradox

Uniform random variables on $[0, 1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1]. \end{cases}$

Uniform random variables on $[0, 1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1]. \end{cases}$
- Then for any $0 \leq a \leq b \leq 1$ we have $P\{X \in [a, b]\} = b - a$.

Uniform random variables on $[0, 1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & x \notin [0, 1]. \end{cases}$
- Then for any $0 \leq a \leq b \leq 1$ we have $P\{X \in [a, b]\} = b - a$.
- Intuition: all locations along the interval $[0, 1]$ equally likely.

Uniform random variables on $[0, 1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1]. \end{cases}$
- Then for any $0 \le a \le b \le 1$ we have $P\{X \in [a,b]\} = b - a$.
- Intuition: all locations along the interval $[0, 1]$ equally likely.
- Say that X is a **uniform random variable on** $[0, 1]$ or that X is **sampled uniformly from** $[0, 1]$.

Properties of uniform random variable on $[0,1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1], \end{cases}$ which implies

$$F_X(a) = \begin{cases} 0 & a < 0 \\ a & a \in [0,1] \\ 1 & a > 1 \end{cases}.$$

Properties of uniform random variable on $[0,1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1], \end{cases}$ which implies

$$F_X(a) = \begin{cases} 0 & a < 0 \\ a & a \in [0,1] \\ 1 & a > 1 \end{cases}.$$

- What is $E[X]$?

Properties of uniform random variable on $[0, 1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1], \end{cases}$ which implies
$$F_X(a) = \begin{cases} 0 & a < 0 \\ a & a \in [0,1] \\ 1 & a > 1 \end{cases}.$$
- What is $E[X]$?
- Guess $1/2$ (since $1/2$ is, you know, in the middle).

Properties of uniform random variable on $[0,1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1], \end{cases}$ which implies

$$F_X(a) = \begin{cases} 0 & a < 0 \\ a & a \in [0,1] \\ 1 & a > 1 \end{cases}.$$

- What is $E[X]$?
- Guess $1/2$ (since $1/2$ is, you know, in the middle).
- Indeed, $\int_{-\infty}^{\infty} f(x)xdx = \int_0^1 xdx = \frac{x^2}{2}\Big|_0^1 = 1/2$.

Properties of uniform random variable on $[0,1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1], \end{cases}$ which implies

$$F_X(a) = \begin{cases} 0 & a < 0 \\ a & a \in [0,1] \\ 1 & a > 1 \end{cases}.$$

- What is $E[X]$?
- Guess $1/2$ (since $1/2$ is, you know, in the middle).
- Indeed, $\int_{-\infty}^{\infty} f(x) x dx = \int_0^1 x dx = \frac{x^2}{2}\Big|_0^1 = 1/2$.
- What is the general moment $E[X^k]$ for $k \geq 0$?

Properties of uniform random variable on $[0,1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1], \end{cases}$ which implies
$$F_X(a) = \begin{cases} 0 & a < 0 \\ a & a \in [0,1] \\ 1 & a > 1 \end{cases}.$$
- What is $E[X]$?
- Guess $1/2$ (since $1/2$ is, you know, in the middle).
- Indeed, $\int_{-\infty}^{\infty} f(x) x dx = \int_0^1 x dx = \frac{x^2}{2}\Big|_0^1 = 1/2$.
- What is the general moment $E[X^k]$ for $k \geq 0$?
- Answer: $1/(k+1)$.

Properties of uniform random variable on $[0,1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1], \end{cases}$ which implies

$$F_X(a) = \begin{cases} 0 & a < 0 \\ a & a \in [0,1] \\ 1 & a > 1 \end{cases}.$$

- What is $E[X]$?
- Guess $1/2$ (since $1/2$ is, you know, in the middle).
- Indeed, $\int_{-\infty}^{\infty} f(x) x dx = \int_0^1 x dx = \frac{x^2}{2}\Big|_0^1 = 1/2$.
- What is the general moment $E[X^k]$ for $k \geq 0$?
- Answer: $1/(k+1)$.
- What would you guess the variance is? Expected square of distance from $1/2$?

Properties of uniform random variable on $[0,1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1], \end{cases}$ which implies
$$F_X(a) = \begin{cases} 0 & a < 0 \\ a & a \in [0,1] \\ 1 & a > 1 \end{cases}.$$
- What is $E[X]$?
- Guess $1/2$ (since $1/2$ is, you know, in the middle).
- Indeed, $\int_{-\infty}^{\infty} f(x) x dx = \int_0^1 x dx = \frac{x^2}{2}\Big|_0^1 = 1/2$.
- What is the general moment $E[X^k]$ for $k \geq 0$?
- Answer: $1/(k+1)$.
- What would you guess the variance is? Expected square of distance from $1/2$?
- It's obviously less than $1/4$, but how much less?

Properties of uniform random variable on $[0,1]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1], \end{cases}$ which implies

$$F_X(a) = \begin{cases} 0 & a < 0 \\ a & a \in [0,1] \\ 1 & a > 1 \end{cases}.$$

- What is $E[X]$?
- Guess $1/2$ (since $1/2$ is, you know, in the middle).
- Indeed, $\int_{-\infty}^{\infty} f(x)xdx = \int_0^1 xdx = \frac{x^2}{2}\Big|_0^1 = 1/2$.
- What is the general moment $E[X^k]$ for $k \geq 0$?
- Answer: $1/(k+1)$.
- What would you guess the variance is? Expected square of distance from $1/2$?
- It's obviously less than $1/4$, but how much less?
- $\mathrm{Var} E[X^2] - E[X]^2 = 1/3 - 1/4 = 1/12$.

Outline

Continuous random variables

Expectation and variance of continuous random variables

Uniform random variable on $[0, 1]$

Uniform random variable on $[\alpha, \beta]$

Measurable sets and a famous paradox

Outline

Uniform random variables on $[\alpha, \beta]$

- Fix $\alpha < \beta$ and suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$

Uniform random variables on $[\alpha, \beta]$

- Fix $\alpha < \beta$ and suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$
- Then for any $\alpha \le a \le b \le \beta$ we have $P\{X \in [a, b]\} = \frac{b-a}{\beta-\alpha}$.

Uniform random variables on $[\alpha, \beta]$

- Fix $\alpha < \beta$ and suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$
- Then for any $\alpha \le a \le b \le \beta$ we have $P\{X \in [a, b]\} = \frac{b-a}{\beta-\alpha}$.
- Intuition: all locations along the interval $[\alpha, \beta]$ are equally likely.

Uniform random variables on $[\alpha, \beta]$

- Fix $\alpha < \beta$ and suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta - \alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$
- Then for any $\alpha \le a \le b \le \beta$ we have $P\{X \in [a, b]\} = \frac{b-a}{\beta-\alpha}$.
- Intuition: all locations along the interval $[\alpha, \beta]$ are equally likely.
- Say that X is a **uniform random variable on** $[\alpha, \beta]$ or that X is **sampled uniformly from** $[\alpha, \beta]$.

Uniform random variables on $[\alpha, \beta]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$

Uniform random variables on $[\alpha, \beta]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta - \alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$
- What is $E[X]$?

Uniform random variables on $[\alpha, \beta]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$
- What is $E[X]$?
- Intuitively, we'd guess the midpoint $\frac{\alpha+\beta}{2}$.

Uniform random variables on $[\alpha, \beta]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$
- What is $E[X]$?
- Intuitively, we'd guess the midpoint $\frac{\alpha+\beta}{2}$.
- What's the cleanest way to prove this?

Uniform random variables on $[\alpha, \beta]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha,\beta] \\ 0 & x \notin [\alpha,\beta]. \end{cases}$
- What is $E[X]$?
- Intuitively, we'd guess the midpoint $\frac{\alpha+\beta}{2}$.
- What's the cleanest way to prove this?
- One approach: let Y be uniform on $[0,1]$ and try to show that $X = (\beta-\alpha)Y + \alpha$ is uniform on $[\alpha,\beta]$.

Uniform random variables on $[\alpha, \beta]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$
- What is $E[X]$?
- Intuitively, we'd guess the midpoint $\frac{\alpha+\beta}{2}$.
- What's the cleanest way to prove this?
- One approach: let Y be uniform on $[0, 1]$ and try to show that $X = (\beta - \alpha)Y + \alpha$ is uniform on $[\alpha, \beta]$.
- Then expectation linearity gives $E[X] = (\beta - \alpha)E[Y] + \alpha = (1/2)(\beta - \alpha) + \alpha = \frac{\alpha+\beta}{2}$.

Uniform random variables on $[\alpha, \beta]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$
- What is $E[X]$?
- Intuitively, we'd guess the midpoint $\frac{\alpha+\beta}{2}$.
- What's the cleanest way to prove this?
- One approach: let Y be uniform on $[0, 1]$ and try to show that $X = (\beta - \alpha)Y + \alpha$ is uniform on $[\alpha, \beta]$.
- Then expectation linearity gives $E[X] = (\beta - \alpha)E[Y] + \alpha = (1/2)(\beta - \alpha) + \alpha = \frac{\alpha+\beta}{2}$.
- Using similar logic, what is the variance $\mathrm{Var}[X]$?

Uniform random variables on $[\alpha, \beta]$

- Suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$
- What is $E[X]$?
- Intuitively, we'd guess the midpoint $\frac{\alpha+\beta}{2}$.
- What's the cleanest way to prove this?
- One approach: let Y be uniform on $[0, 1]$ and try to show that $X = (\beta - \alpha)Y + \alpha$ is uniform on $[\alpha, \beta]$.
- Then expectation linearity gives $E[X] = (\beta - \alpha)E[Y] + \alpha = (1/2)(\beta - \alpha) + \alpha = \frac{\alpha+\beta}{2}$.
- Using similar logic, what is the variance $\mathrm{Var}[X]$?
- Answer: $\mathrm{Var}[X] = \mathrm{Var}[(\beta - \alpha)Y + \alpha] = \mathrm{Var}[(\beta - \alpha)Y] = (\beta - \alpha)^2\mathrm{Var}[Y] = (\beta - \alpha)^2/12$.

Outline

Continuous random variables

Expectation and variance of continuous random variables

Uniform random variable on $[0, 1]$

Uniform random variable on $[\alpha, \beta]$

Measurable sets and a famous paradox

Outline

Continuous random variables

Expectation and variance of continuous random variables

Uniform random variable on $[0, 1]$

Uniform random variable on $[\alpha, \beta]$

Measurable sets and a famous paradox

Uniform measure: is probability defined for all subsets?

- One of the very simplest probability density functions is
  $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & 0 \notin [0,1]. \end{cases}.$

Uniform measure: is probability defined for all subsets?

- One of the very simplest probability density functions is $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & 0 \notin [0,1]. \end{cases}$.
- If $B \subset [0,1]$ is an interval, then $P\{X \in B\}$ is the length of that interval.

Uniform measure: is probability defined for all subsets?

- One of the very simplest probability density functions is $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & 0 \not\in [0,1]. \end{cases}$.
- If $B \subset [0,1]$ is an interval, then $P\{X \in B\}$ is the length of that interval.
- Generally, if $B \subset [0,1]$ then $P\{X \in B\} = \int_B 1 dx = \int 1_B(x) dx$ is the "total volume" or "total length" of the set B.

Uniform measure: is probability defined for all subsets?

- One of the very simplest probability density functions is $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & 0 \notin [0,1]. \end{cases}$.
- If $B \subset [0,1]$ is an interval, then $P\{X \in B\}$ is the length of that interval.
- Generally, if $B \subset [0,1]$ then $P\{X \in B\} = \int_B 1 dx = \int 1_B(x) dx$ is the "total volume" or "total length" of the set B.
- What if B is the set of all rational numbers?

Uniform measure: is probability defined for all subsets?

- One of the very simplest probability density functions is $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & 0 \notin [0,1]. \end{cases}$.
- If $B \subset [0,1]$ is an interval, then $P\{X \in B\}$ is the length of that interval.
- Generally, if $B \subset [0,1]$ then $P\{X \in B\} = \int_B 1 dx = \int 1_B(x) dx$ is the "total volume" or "total length" of the set B.
- What if B is the set of all rational numbers?
- How do we mathematically define the volume of an arbitrary set B?

Idea behind parodox

- **Hypothetical:** Consider the interval $[0, 1)$ with the two endpoints glued together (so it looks like a circle). *What if* we could partition $[0, 1)$ into a countably infinite collection of disjoint sets that all looked the same (up to a rotation of the circle) and thus had to have the same probability?

Idea behind parodox

- **Hypothetical:** Consider the interval $[0, 1)$ with the two endpoints glued together (so it looks like a circle). *What if* we could partition $[0, 1)$ into a countably infinite collection of disjoint sets that all looked the same (up to a rotation of the circle) and thus had to have the same probability?
- If that probability was zero, then (by countable additivity) probability of whole circle would be zero, a contradiction.

Idea behind parodox

- **Hypothetical:** Consider the interval $[0, 1)$ with the two endpoints glued together (so it looks like a circle). *What if* we could partition $[0, 1)$ into a countably infinite collection of disjoint sets that all looked the same (up to a rotation of the circle) and thus had to have the same probability?
- If that probability was zero, then (by countable additivity) probability of whole circle would be zero, a contradiction.
- But if that probability were a number greater than zero the probability of whole circle would be infinite, also a contradiction...

Idea behind parodox

- **Hypothetical:** Consider the interval $[0,1)$ with the two endpoints glued together (so it looks like a circle). *What if* we could partition $[0,1)$ into a countably infinite collection of disjoint sets that all looked the same (up to a rotation of the circle) and thus had to have the same probability?
- If that probability was zero, then (by countable additivity) probability of whole circle would be zero, a contradiction.
- But if that probability were a number greater than zero the probability of whole circle would be infinite, also a contradiction...
- **Related problem:** *if* (in a non-atomic world, where mass was infinitely divisible) you could cut a cake into countably infinitely many pieces all of the same weight, how much would each piece weigh?

Idea behind parodox

- **Hypothetical:** Consider the interval $[0, 1)$ with the two endpoints glued together (so it looks like a circle). *What if* we could partition $[0, 1)$ into a countably infinite collection of disjoint sets that all looked the same (up to a rotation of the circle) and thus had to have the same probability?
- If that probability was zero, then (by countable additivity) probability of whole circle would be zero, a contradiction.
- But if that probability were a number greater than zero the probability of whole circle would be infinite, also a contradiction...
- **Related problem:** *if* (in a non-atomic world, where mass was infinitely divisible) you could cut a cake into countably infinitely many pieces all of the same weight, how much would each piece weigh?
- **Question:** Is it really possible to partition $[0, 1)$ into countably many identical (up to rotation) pieces?

Cutting things into identical slices: a warmup problem

- **Consider** the set of numbers $\{0, 1, 2, \ldots, 99\}$.

Cutting things into identical slices: a warmup problem

- **Consider** the set of numbers $\{0, 1, 2, \ldots, 99\}$.
- **Let's suggest** one fancy way to divide this set into ten equal subsets that are translations of each other modulo 100.

Cutting things into identical slices: a warmup problem

- **Consider** the set of numbers $\{0, 1, 2, \ldots, 99\}$.
- **Let's suggest** one fancy way to divide this set into ten equal subsets that are translations of each other modulo 100.
- Two numbers are **equivalent modulo** 10 if their difference is a multiple of 10 (so they end in same digit). Pick a set $S \subset \{0, 1, 2, \ldots, 99\}$ with one number from each *equivalence class*, e.g., $S = \{40, 21, 42, 53, 94, 5, 76, 27, 28, 39\}$.

Cutting things into identical slices: a warmup problem

- **Consider** the set of numbers $\{0, 1, 2, \ldots, 99\}$.
- **Let's suggest** one fancy way to divide this set into ten equal subsets that are translations of each other modulo 100.
- Two numbers are **equivalent modulo** 10 if their difference is a multiple of 10 (so they end in same digit). Pick a set $S \subset \{0, 1, 2, \ldots, 99\}$ with one number from each *equivalence class*, e.g., $S = \{40, 21, 42, 53, 94, 5, 76, 27, 28, 39\}$.
- Then for each $j \in \{0, 10, 20, \ldots, 90\}$ define the set $S_j = \{s + j : s \in S\}$, where addition is modulo 100.

Cutting things into identical slices: a warmup problem

- **Consider** the set of numbers $\{0, 1, 2, \dots, 99\}$.
- **Let's suggest** one fancy way to divide this set into ten equal subsets that are translations of each other modulo 100.
- Two numbers are **equivalent modulo** 10 if their difference is a multiple of 10 (so they end in same digit). Pick a set $S \subset \{0, 1, 2, \dots, 99\}$ with one number from each *equivalence class*, e.g., $S = \{40, 21, 42, 53, 94, 5, 76, 27, 28, 39\}$.
- Then for each $j \in \{0, 10, 20, \dots, 90\}$ define the set $S_j = \{s + j : s \in S\}$, where addition is modulo 100.
- Now observe that every number in $\{0, 1, 2, \dots, 99\}$ lies in exactly one of the ten S_j sets we have defined.

Cutting things into identical slices: a warmup problem

- **Consider** the set of numbers $\{0, 1, 2, \ldots, 99\}$.
- **Let's suggest** one fancy way to divide this set into ten equal subsets that are translations of each other modulo 100.
- Two numbers are **equivalent modulo** 10 if their difference is a multiple of 10 (so they end in same digit). Pick a set $S \subset \{0, 1, 2, \ldots, 99\}$ with one number from each *equivalence class*, e.g., $S = \{40, 21, 42, 53, 94, 5, 76, 27, 28, 39\}$.
- Then for each $j \in \{0, 10, 20, \ldots, 90\}$ define the set $S_j = \{s + j : s \in S\}$, where addition is modulo 100.
- Now observe that every number in $\{0, 1, 2, \ldots, 99\}$ lies in exactly one of the ten S_j sets we have defined.
- On next slide, we're going to do something similar with $[0, 1)$ in place of $\{0, 1, 2, \ldots, 99\}$ and the **rational numbers in** $[0, 1)$ in place of $\{0, 10, 20, \ldots, 90\}$.

Formulating the paradox precisely

- Consider **wrap-around translations** $\tau_r(x) = (x + r) \mod 1$.

Formulating the paradox precisely

- Consider **wrap-around translations** $\tau_r(x) = (x + r) \mod 1$.
- We expect $\tau_r(B)$ to have same probability as B.

Formulating the paradox precisely

- Consider **wrap-around translations** $\tau_r(x) = (x + r) \mod 1$.
- We expect $\tau_r(B)$ to have same probability as B.
- Call x, y "equivalent modulo rationals" if $x - y$ is rational (e.g., $x = \pi - 3$ and $y = \pi - 9/4$). An **equivalence class** is the set of points in $[0, 1)$ equivalent to some given point.

Formulating the paradox precisely

- Consider **wrap-around translations** $\tau_r(x) = (x + r) \mod 1$.
- We expect $\tau_r(B)$ to have same probability as B.
- Call x, y "equivalent modulo rationals" if $x - y$ is rational (e.g., $x = \pi - 3$ and $y = \pi - 9/4$). An **equivalence class** is the set of points in $[0, 1)$ equivalent to some given point.
- There are uncountably many of these classes.

Formulating the paradox precisely

- Consider **wrap-around translations** $\tau_r(x) = (x + r) \mod 1$.
- We expect $\tau_r(B)$ to have same probability as B.
- Call x, y "equivalent modulo rationals" if $x - y$ is rational (e.g., $x = \pi - 3$ and $y = \pi - 9/4$). An **equivalence class** is the set of points in $[0, 1)$ equivalent to some given point.
- There are uncountably many of these classes.
- Let $A \subset [0, 1)$ contain **one** point from each class. For each $x \in [0, 1)$, there is **one** $a \in A$ such that $r = x - a$ is rational.

Formulating the paradox precisely

- Consider **wrap-around translations** $\tau_r(x) = (x + r) \mod 1$.
- We expect $\tau_r(B)$ to have same probability as B.
- Call x, y "equivalent modulo rationals" if $x - y$ is rational (e.g., $x = \pi - 3$ and $y = \pi - 9/4$). An **equivalence class** is the set of points in $[0, 1)$ equivalent to some given point.
- There are uncountably many of these classes.
- Let $A \subset [0, 1)$ contain **one** point from each class. For each $x \in [0, 1)$, there is **one** $a \in A$ such that $r = x - a$ is rational.
- Then each x in $[0, 1)$ lies in $\tau_r(A)$ for **one** rational $r \in [0, 1)$.

Formulating the paradox precisely

- Consider **wrap-around translations** $\tau_r(x) = (x + r) \mod 1$.
- We expect $\tau_r(B)$ to have same probability as B.
- Call x, y "equivalent modulo rationals" if $x - y$ is rational (e.g., $x = \pi - 3$ and $y = \pi - 9/4$). An **equivalence class** is the set of points in $[0, 1)$ equivalent to some given point.
- There are uncountably many of these classes.
- Let $A \subset [0, 1)$ contain **one** point from each class. For each $x \in [0, 1)$, there is **one** $a \in A$ such that $r = x - a$ is rational.
- Then each x in $[0, 1)$ lies in $\tau_r(A)$ for **one** rational $r \in [0, 1)$.
- Thus $[0, 1) = \cup \tau_r(A)$ as r ranges over rationals in $[0, 1)$.

Formulating the paradox precisely

- Consider **wrap-around translations** $\tau_r(x) = (x + r) \mod 1$.
- We expect $\tau_r(B)$ to have same probability as B.
- Call x, y "equivalent modulo rationals" if $x - y$ is rational (e.g., $x = \pi - 3$ and $y = \pi - 9/4$). An **equivalence class** is the set of points in $[0, 1)$ equivalent to some given point.
- There are uncountably many of these classes.
- Let $A \subset [0, 1)$ contain **one** point from each class. For each $x \in [0, 1)$, there is **one** $a \in A$ such that $r = x - a$ is rational.
- Then each x in $[0, 1)$ lies in $\tau_r(A)$ for **one** rational $r \in [0, 1)$.
- Thus $[0, 1) = \cup \tau_r(A)$ as r ranges over rationals in $[0, 1)$.
- If $P(A) = 0$, then $P(S) = \sum_r P(\tau_r(A)) = 0$. If $P(A) > 0$ then $P(S) = \sum_r P(\tau_r(A)) = \infty$. Contradicts $P(S) = 1$ axiom.

Three ways to get around this

- 1. **Re-examine axioms of mathematics:** the very *existence* of a set A with one element from each equivalence class is consequence of so-called **axiom of choice**. Removing that axiom makes paradox goes away, since one can just suppose (pretend?) these kinds of sets don't exist.

Three ways to get around this

- 1. **Re-examine axioms of mathematics:** the very *existence* of a set A with one element from each equivalence class is consequence of so-called **axiom of choice**. Removing that axiom makes paradox goes away, since one can just suppose (pretend?) these kinds of sets don't exist.
- 2. **Re-examine axioms of probability:** Replace *countable additivity* with *finite additivity*? (Doesn't fully solve problem: look up Banach-Tarski.)

Three ways to get around this

- 1. **Re-examine axioms of mathematics:** the very *existence* of a set A with one element from each equivalence class is consequence of so-called **axiom of choice**. Removing that axiom makes paradox goes away, since one can just suppose (pretend?) these kinds of sets don't exist.
- 2. **Re-examine axioms of probability:** Replace *countable additivity* with *finite additivity*? (Doesn't fully solve problem: look up Banach-Tarski.)
- 3. **Keep the axiom of choice and countable additivity but don't define probabilities of all sets:** Instead of defining $P(B)$ for *every* subset B of sample space, restrict attention to a family of so-called **"measurable"** sets.

Three ways to get around this

- 1. **Re-examine axioms of mathematics:** the very *existence* of a set A with one element from each equivalence class is consequence of so-called **axiom of choice**. Removing that axiom makes paradox goes away, since one can just suppose (pretend?) these kinds of sets don't exist.
- 2. **Re-examine axioms of probability:** Replace *countable additivity* with *finite additivity*? (Doesn't fully solve problem: look up Banach-Tarski.)
- 3. **Keep the axiom of choice and countable additivity but don't define probabilities of all sets:** Instead of defining $P(B)$ for *every* subset B of sample space, restrict attention to a family of so-called **"measurable"** sets.
- Most mainstream probability and analysis takes the third approach.

Three ways to get around this

- 1. **Re-examine axioms of mathematics:** the very *existence* of a set A with one element from each equivalence class is consequence of so-called **axiom of choice**. Removing that axiom makes paradox goes away, since one can just suppose (pretend?) these kinds of sets don't exist.
- 2. **Re-examine axioms of probability:** Replace *countable additivity* with *finite additivity*? (Doesn't fully solve problem: look up Banach-Tarski.)
- 3. **Keep the axiom of choice and countable additivity but don't define probabilities of all sets:** Instead of defining $P(B)$ for *every* subset B of sample space, restrict attention to a family of so-called **"measurable"** sets.
- Most mainstream probability and analysis takes the third approach.
- In practice, sets we care about (e.g., countable unions of points and intervals) tend to be measurable.

Perspective

- More advanced courses in probability and analysis (such as 18.125 and 18.175) spend a significant amount of time rigorously constructing a class of so-called **measurable sets** and the so-called **Lebesgue measure**, which assigns a real number (a measure) to each of these sets.

Perspective

- More advanced courses in probability and analysis (such as 18.125 and 18.175) spend a significant amount of time rigorously constructing a class of so-called **measurable sets** and the so-called **Lebesgue measure**, which assigns a real number (a measure) to each of these sets.
- These courses also replace the **Riemann integral** with the so-called **Lebesgue integral**.

Perspective

- More advanced courses in probability and analysis (such as 18.125 and 18.175) spend a significant amount of time rigorously constructing a class of so-called **measurable sets** and the so-called **Lebesgue measure**, which assigns a real number (a measure) to each of these sets.
- These courses also replace the **Riemann integral** with the so-called **Lebesgue integral**.
- We will not treat these topics any further in this course.

Perspective

- More advanced courses in probability and analysis (such as 18.125 and 18.175) spend a significant amount of time rigorously constructing a class of so-called **measurable sets** and the so-called **Lebesgue measure**, which assigns a real number (a measure) to each of these sets.
- These courses also replace the **Riemann integral** with the so-called **Lebesgue integral**.
- We will not treat these topics any further in this course.
- We usually limit our attention to probability density functions f and sets B for which the ordinary Riemann integral $\int 1_B(x) f(x) dx$ is well defined.

Perspective

- More advanced courses in probability and analysis (such as 18.125 and 18.175) spend a significant amount of time rigorously constructing a class of so-called **measurable sets** and the so-called **Lebesgue measure**, which assigns a real number (a measure) to each of these sets.
- These courses also replace the **Riemann integral** with the so-called **Lebesgue integral**.
- We will not treat these topics any further in this course.
- We usually limit our attention to probability density functions f and sets B for which the ordinary Riemann integral $\int 1_B(x)f(x)dx$ is well defined.
- Riemann integration is a mathematically rigorous theory. It's just not as robust as Lebesgue integration.