18.440: Lecture 21

More continuous random variables

Scott Sheffield

MIT

Outline

Outline

Gamma distribution

Cauchy distribution

Beta distribution

Defining gamma function Γ

- Last time we found that if X is geometric with rate 1 and $n \geq 0$ then $E[X^n] = \int_0^\infty x^n e^{-x} dx = n!$.

Defining gamma function Γ

- Last time we found that if X is geometric with rate 1 and $n \geq 0$ then $E[X^n] = \int_0^\infty x^n e^{-x} dx = n!$.
- This expectation $E[X^n]$ is actually well defined whenever $n > -1$. Set $\alpha = n + 1$. The following quantity is well defined for any $\alpha > 0$:
  $\Gamma(\alpha) := E[X^{\alpha-1}] = \int_0^\infty x^{\alpha-1} e^{-x} dx = (\alpha - 1)!$.

Defining gamma function Γ

- Last time we found that if X is geometric with rate 1 and $n \geq 0$ then $E[X^n] = \int_0^\infty x^n e^{-x} dx = n!$.
- This expectation $E[X^n]$ is actually well defined whenever $n > -1$. Set $\alpha = n + 1$. The following quantity is well defined for any $\alpha > 0$:
  $\Gamma(\alpha) := E[X^{\alpha-1}] = \int_0^\infty x^{\alpha-1} e^{-x} dx = (\alpha - 1)!$.
- So $\Gamma(\alpha)$ extends the function $(\alpha - 1)!$ (as defined for *strictly positive* integers α) to the positive reals.

Defining gamma function Γ

- Last time we found that if X is geometric with rate 1 and $n \geq 0$ then $E[X^n] = \int_0^\infty x^n e^{-x} dx = n!$.
- This expectation $E[X^n]$ is actually well defined whenever $n > -1$. Set $\alpha = n+1$. The following quantity is well defined for any $\alpha > 0$:
  $\Gamma(\alpha) := E[X^{\alpha-1}] = \int_0^\infty x^{\alpha-1} e^{-x} dx = (\alpha-1)!$.
- So $\Gamma(\alpha)$ extends the function $(\alpha-1)!$ (as defined for *strictly positive* integers α) to the positive reals.
- Vexing notational issue: why define Γ so that $\Gamma(\alpha) = (\alpha-1)!$ instead of $\Gamma(\alpha) = \alpha!$?

Defining gamma function Γ

- Last time we found that if X is geometric with rate 1 and $n \geq 0$ then $E[X^n] = \int_0^\infty x^n e^{-x} dx = n!$.
- This expectation $E[X^n]$ is actually well defined whenever $n > -1$. Set $\alpha = n + 1$. The following quantity is well defined for any $\alpha > 0$:
  $\Gamma(\alpha) := E[X^{\alpha-1}] = \int_0^\infty x^{\alpha-1} e^{-x} dx = (\alpha - 1)!$.
- So $\Gamma(\alpha)$ extends the function $(\alpha - 1)!$ (as defined for *strictly positive* integers α) to the positive reals.
- Vexing notational issue: why define Γ so that $\Gamma(\alpha) = (\alpha - 1)!$ instead of $\Gamma(\alpha) = \alpha!$?
- At least it's kind of convenient that Γ is defined on $(0, \infty)$ instead of $(-1, \infty)$.

Recall: geometric and negative binomials

- The sum X of n independent geometric random variables of parameter p is negative binomial with parameter (n, p).

Recall: geometric and negative binomials

- The sum X of n independent geometric random variables of parameter p is negative binomial with parameter (n, p).
- Waiting for the nth heads. What is $P\{X = k\}$?

Recall: geometric and negative binomials

- The sum X of n independent geometric random variables of parameter p is negative binomial with parameter (n, p).
- Waiting for the nth heads. What is $P\{X = k\}$?
- Answer: $\binom{k-1}{n-1}p^{n-1}(1-p)^{k-n}p$.

Recall: geometric and negative binomials

- The sum X of n independent geometric random variables of parameter p is negative binomial with parameter (n, p).
- Waiting for the nth heads. What is $P\{X = k\}$?
- Answer: $\binom{k-1}{n-1}p^{n-1}(1-p)^{k-n}p$.
- What's the continuous (Poisson point process) version of "waiting for the nth event"?

Poisson point process limit

- Recall that we can approximate a Poisson process of rate λ by tossing N coins per time unit and taking $p = \lambda/N$.

Poisson point process limit

- Recall that we can approximate a Poisson process of rate λ by tossing N coins per time unit and taking $p = \lambda/N$.
- Let's fix a rational number x and try to figure out the probability that that the nth coin toss happens at time x (i.e., on exactly xNth trials, assuming xN is an integer).

Poisson point process limit

- Recall that we can approximate a Poisson process of rate λ by tossing N coins per time unit and taking $p = \lambda/N$.
- Let's fix a rational number x and try to figure out the probability that that the nth coin toss happens at time x (i.e., on exactly xNth trials, assuming xN is an integer).
- Write $p = \lambda/N$ and $k = xN$. (Note $p = \lambda x/k$.)

Poisson point process limit

- Recall that we can approximate a Poisson process of rate λ by tossing N coins per time unit and taking $p = \lambda/N$.
- Let's fix a rational number x and try to figure out the probability that that the nth coin toss happens at time x (i.e., on exactly xNth trials, assuming xN is an integer).
- Write $p = \lambda/N$ and $k = xN$. (Note $p = \lambda x/k$.)
- For large N, $\binom{k-1}{n-1} p^{n-1}(1-p)^{k-n}p$ is

$$\frac{(k-1)(k-2)\ldots(k-n+1)}{(n-1)!} p^{n-1}(1-p)^{k-n}p$$

$$\approx \frac{k^{n-1}}{(n-1)!} p^{n-1} e^{-x\lambda} p = \frac{1}{N}\Big(\frac{(\lambda x)^{(n-1)} e^{-\lambda x}\lambda}{(n-1)!}\Big).$$

Defining Γ distribution

- The probability from previous side, $\frac{1}{N}\left(\frac{(\lambda x)^{(n-1)}e^{-\lambda x}\lambda}{(n-1)!}\right)$ suggests the form for a continuum random variable.

Defining Γ distribution

- The probability from previous side, $\frac{1}{N}\left(\frac{(\lambda x)^{(n-1)}e^{-\lambda x}\lambda}{(n-1)!}\right)$ suggests the form for a continuum random variable.
- Replace n (generally integer valued) with α (which we will eventually allow be to be any real number).

Defining Γ distribution

- The probability from previous side, $\frac{1}{N}\left(\frac{(\lambda x)^{(n-1)}e^{-\lambda x}\lambda}{(n-1)!}\right)$ suggests the form for a continuum random variable.
- Replace n (generally integer valued) with α (which we will eventually allow be to be any real number).
- Say that random variable X has gamma distribution with parameters (α, λ) if $f_X(x) = \begin{cases} \frac{(\lambda x)^{\alpha-1}e^{-\lambda x}\lambda}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$.

Defining Γ distribution

- The probability from previous side, $\frac{1}{N}\left(\frac{(\lambda x)^{(n-1)}e^{-\lambda x}\lambda}{(n-1)!}\right)$ suggests the form for a continuum random variable.
- Replace n (generally integer valued) with α (which we will eventually allow be to be any real number).
- Say that random variable X has gamma distribution with parameters (α, λ) if $f_X(x) = \begin{cases} \frac{(\lambda x)^{\alpha-1}e^{-\lambda x}\lambda}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$.
- Waiting time interpretation makes sense only for integer α, but distribution is defined for general positive α.

Outline

Gamma distribution

Cauchy distribution

Beta distribution

Outline

Gamma distribution

Cauchy distribution

Beta distribution

Cauchy distribution

- A standard **Cauchy random variable** is a random real number with probability density $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.

Cauchy distribution

- A standard **Cauchy random variable** is a random real number with probability density $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.
- There is a "spinning flashlight" interpretation. Put a flashlight at $(0, 1)$, spin it to a uniformly random angle in $[-\pi/2, \pi/2]$, and consider point X where light beam hits the x-axis.

Cauchy distribution

- A standard **Cauchy random variable** is a random real number with probability density $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.
- There is a "spinning flashlight" interpretation. Put a flashlight at $(0,1)$, spin it to a uniformly random angle in $[-\pi/2, \pi/2]$, and consider point X where light beam hits the x-axis.
- $F_X(x) = P\{X \le x\} = P\{\tan\theta \le x\} = P\{\theta \le tan^{-1}x\} = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}x$.

Cauchy distribution

- A standard **Cauchy random variable** is a random real number with probability density $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.
- There is a "spinning flashlight" interpretation. Put a flashlight at $(0,1)$, spin it to a uniformly random angle in $[-\pi/2, \pi/2]$, and consider point X where light beam hits the x-axis.
- $F_X(x) = P\{X \leq x\} = P\{\tan\theta \leq x\} = P\{\theta \leq tan^{-1}x\} = \frac{1}{2} + \frac{1}{\pi}\tan^{-1} x$.
- Find $f_X(x) = \frac{d}{dx}F(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.

Cauchy distribution: Brownian motion interpretation

- The light beam travels in (randomly directed) straight line. There's a windier random path called Brownian motion.

Cauchy distribution: Brownian motion interpretation

- The light beam travels in (randomly directed) straight line. There's a windier random path called Brownian motion.
- If you do a simple random walk on a grid and take the grid size to zero, then you get Brownian motion as a limit.

Cauchy distribution: Brownian motion interpretation

- The light beam travels in (randomly directed) straight line. There's a windier random path called Brownian motion.
- If you do a simple random walk on a grid and take the grid size to zero, then you get Brownian motion as a limit.
- We will not give a complete mathematical description of Brownian motion here, just one nice fact.

Cauchy distribution: Brownian motion interpretation

- The light beam travels in (randomly directed) straight line. There's a windier random path called Brownian motion.
- If you do a simple random walk on a grid and take the grid size to zero, then you get Brownian motion as a limit.
- We will not give a complete mathematical description of Brownian motion here, just one nice fact.
- FACT: start Brownian motion at point (x, y) in the upper half plane. Probability it hits negative x-axis before positive x-axis is $\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\frac{y}{x}$. Linear function of angle between positive x-axis and line through $(0, 0)$ and (x, y).

Cauchy distribution: Brownian motion interpretation

- The light beam travels in (randomly directed) straight line. There's a windier random path called Brownian motion.
- If you do a simple random walk on a grid and take the grid size to zero, then you get Brownian motion as a limit.
- We will not give a complete mathematical description of Brownian motion here, just one nice fact.
- FACT: start Brownian motion at point (x, y) in the upper half plane. Probability it hits negative x-axis before positive x-axis is $\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\frac{y}{x}$. Linear function of angle between positive x-axis and line through $(0,0)$ and (x, y).
- Start Brownian motion at $(0,1)$ and let X be the location of the first point on the x-axis it hits. What's $P\{X < a\}$?

Cauchy distribution: Brownian motion interpretation

- The light beam travels in (randomly directed) straight line. There's a windier random path called Brownian motion.
- If you do a simple random walk on a grid and take the grid size to zero, then you get Brownian motion as a limit.
- We will not give a complete mathematical description of Brownian motion here, just one nice fact.
- FACT: start Brownian motion at point (x, y) in the upper half plane. Probability it hits negative x-axis before positive x-axis is $\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\frac{y}{x}$. Linear function of angle between positive x-axis and line through $(0, 0)$ and (x, y).
- Start Brownian motion at $(0, 1)$ and let X be the location of the first point on the x-axis it hits. What's $P\{X < a\}$?
- Applying FACT, translation invariance, reflection symmetry: $P\{X < x\} = P\{X > -x\} = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\frac{1}{x}$.

Cauchy distribution: Brownian motion interpretation

- The light beam travels in (randomly directed) straight line. There's a windier random path called Brownian motion.
- If you do a simple random walk on a grid and take the grid size to zero, then you get Brownian motion as a limit.
- We will not give a complete mathematical description of Brownian motion here, just one nice fact.
- FACT: start Brownian motion at point (x, y) in the upper half plane. Probability it hits negative x-axis before positive x-axis is $\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\frac{y}{x}$. Linear function of angle between positive x-axis and line through $(0, 0)$ and (x, y).
- Start Brownian motion at $(0, 1)$ and let X be the location of the first point on the x-axis it hits. What's $P\{X < a\}$?
- Applying FACT, translation invariance, reflection symmetry: $P\{X < x\} = P\{X > -x\} = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\frac{1}{x}$.
- So X is a standard Cauchy random variable.

Question: what if we start at $(0, 2)$?

- Start at $(0, 2)$. Let Y be first point on x-axis hit by Brownian motion. Again, same probability distribution as point hit by flashlight trajectory.

Question: what if we start at $(0,2)$?

- Start at $(0,2)$. Let Y be first point on x-axis hit by Brownian motion. Again, same probability distribution as point hit by flashlight trajectory.
- Flashlight point of view: Y has the same law as $2X$ where X is standard Cauchy.

Question: what if we start at $(0, 2)$?

- Start at $(0, 2)$. Let Y be first point on x-axis hit by Brownian motion. Again, same probability distribution as point hit by flashlight trajectory.
- Flashlight point of view: Y has the same law as $2X$ where X is standard Cauchy.
- Brownian point of view: Y has same law as $X_1 + X_2$ where X_1 and X_2 are standard Cauchy.

Question: what if we start at $(0,2)$?

- Start at $(0,2)$. Let Y be first point on x-axis hit by Brownian motion. Again, same probability distribution as point hit by flashlight trajectory.
- Flashlight point of view: Y has the same law as $2X$ where X is standard Cauchy.
- Brownian point of view: Y has same law as $X_1 + X_2$ where X_1 and X_2 are standard Cauchy.
- But wait a minute. $\mathrm{Var}(Y) = 4\mathrm{Var}(X)$ and by independence $\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 2\mathrm{Var}(X_2)$. Can this be right?

Question: what if we start at $(0, 2)$?

- Start at $(0, 2)$. Let Y be first point on x-axis hit by Brownian motion. Again, same probability distribution as point hit by flashlight trajectory.
- Flashlight point of view: Y has the same law as $2X$ where X is standard Cauchy.
- Brownian point of view: Y has same law as $X_1 + X_2$ where X_1 and X_2 are standard Cauchy.
- But wait a minute. $\mathrm{Var}(Y) = 4\mathrm{Var}(X)$ and by independence $\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 2\mathrm{Var}(X_2)$. Can this be right?
- Cauchy distribution doesn't have finite variance or mean.

Question: what if we start at $(0, 2)$?

- Start at $(0, 2)$. Let Y be first point on x-axis hit by Brownian motion. Again, same probability distribution as point hit by flashlight trajectory.
- Flashlight point of view: Y has the same law as $2X$ where X is standard Cauchy.
- Brownian point of view: Y has same law as $X_1 + X_2$ where X_1 and X_2 are standard Cauchy.
- But wait a minute. $\mathrm{Var}(Y) = 4\mathrm{Var}(X)$ and by independence $\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 2\mathrm{Var}(X_2)$. Can this be right?
- Cauchy distribution doesn't have finite variance or mean.
- Some standard facts we'll learn later in the course (central limit theorem, law of large numbers) don't apply to it.

Outline

Gamma distribution

Cauchy distribution

Beta distribution

Outline

Gamma distribution

Cauchy distribution

Beta distribution

Beta distribution: Alice and Bob revisited

- Suppose I have a coin with a heads probability p that I don't know much about.

Beta distribution: Alice and Bob revisited

- Suppose I have a coin with a heads probability p that I don't know much about.
- What do I mean by not knowing anything? Let's say that I think p is equally likely to be any of the numbers $\{0, .1, .2, .3, .4, \dots, .9, 1\}$.

Beta distribution: Alice and Bob revisited

- Suppose I have a coin with a heads probability p that I don't know much about.
- What do I mean by not knowing anything? Let's say that I think p is equally likely to be any of the numbers $\{0, .1, .2, .3, .4, \ldots, .9, 1\}$.
- Now imagine a multi-stage experiment where I first choose p and then I toss n coins.

Beta distribution: Alice and Bob revisited

- Suppose I have a coin with a heads probability p that I don't know much about.
- What do I mean by not knowing anything? Let's say that I think p is equally likely to be any of the numbers $\{0, .1, .2, .3, .4, \ldots, .9, 1\}$.
- Now imagine a multi-stage experiment where I first choose p and then I toss n coins.
- Given that number h of heads is $a-1$, and $b-1$ tails, what's *conditional* probability p was a certain value x?

Beta distribution: Alice and Bob revisited

- Suppose I have a coin with a heads probability p that I don't know much about.
- What do I mean by not knowing anything? Let's say that I think p is equally likely to be any of the numbers $\{0, .1, .2, .3, .4, \ldots, .9, 1\}$.
- Now imagine a multi-stage experiment where I first choose p and then I toss n coins.
- Given that number h of heads is $a-1$, and $b-1$ tails, what's *conditional* probability p was a certain value x?
- $P\Big(p = x | h = (a-1)\Big) = \frac{\frac{1}{11}\binom{n}{a-1}x^{a-1}(1-x)^{b-1}}{P\{h=(a-1)\}}$ which is $x^{a-1}(1-x)^{b-1}$ times a constant that doesn't depend on x.

Beta distribution

- Suppose I have a coin with a heads probability p that I *really* don't know anything about. Let's say p is uniform on $[0, 1]$.

Beta distribution

- Suppose I have a coin with a heads probability p that I *really* don't know anything about. Let's say p is uniform on $[0, 1]$.
- Now imagine a multi-stage experiment where I first choose p uniformly from $[0, 1]$ and then I toss n coins.

Beta distribution

- Suppose I have a coin with a heads probability p that I *really* don't know anything about. Let's say p is uniform on $[0, 1]$.
- Now imagine a multi-stage experiment where I first choose p uniformly from $[0, 1]$ and then I toss n coins.
- If I get, say, $a - 1$ heads and $b - 1$ tails, then what is the *conditional* probability density for p?

Beta distribution

- Suppose I have a coin with a heads probability p that I *really* don't know anything about. Let's say p is uniform on $[0, 1]$.
- Now imagine a multi-stage experiment where I first choose p uniformly from $[0, 1]$ and then I toss n coins.
- If I get, say, $a - 1$ heads and $b - 1$ tails, then what is the *conditional* probability density for p?
- Turns out to be a constant (that doesn't depend on x) times $x^{a-1}(1 - x)^{b-1}$.

Beta distribution

- Suppose I have a coin with a heads probability p that I *really* don't know anything about. Let's say p is uniform on $[0,1]$.
- Now imagine a multi-stage experiment where I first choose p uniformly from $[0,1]$ and then I toss n coins.
- If I get, say, $a-1$ heads and $b-1$ tails, then what is the *conditional* probability density for p?
- Turns out to be a constant (that doesn't depend on x) times $x^{a-1}(1-x)^{b-1}$.
- $\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}$ on $[0,1]$, where $B(a,b)$ is constant chosen to make integral one. Can be shown that $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Beta distribution

- Suppose I have a coin with a heads probability p that I *really* don't know anything about. Let's say p is uniform on $[0,1]$.
- Now imagine a multi-stage experiment where I first choose p uniformly from $[0,1]$ and then I toss n coins.
- If I get, say, $a-1$ heads and $b-1$ tails, then what is the *conditional* probability density for p?
- Turns out to be a constant (that doesn't depend on x) times $x^{a-1}(1-x)^{b-1}$.
- $\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}$ on $[0,1]$, where $B(a,b)$ is constant chosen to make integral one. Can be shown that $B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.
- What is $E[X]$?

Beta distribution

- Suppose I have a coin with a heads probability p that I *really* don't know anything about. Let's say p is uniform on $[0, 1]$.
- Now imagine a multi-stage experiment where I first choose p uniformly from $[0, 1]$ and then I toss n coins.
- If I get, say, $a - 1$ heads and $b - 1$ tails, then what is the *conditional* probability density for p?
- Turns out to be a constant (that doesn't depend on x) times $x^{a-1}(1-x)^{b-1}$.
- $\frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$ on $[0, 1]$, where $B(a, b)$ is constant chosen to make integral one. Can be shown that $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.
- What is $E[X]$?
- Answer: $\frac{a}{a+b}$.