BASIC DISCRETE RANDOM VARIABLES \( X \) (using \( q = 1 - p \))

1. **Binomial** \( (n,p) \): \( p_X(k) = \binom{n}{k} p^k q^{n-k} \) and \( E[X] = np \) and \( \text{Var}[X] = npq \).
2. **Poisson** \( \lambda \): \( p_X(k) = e^{-\lambda} \lambda^k / k! \) and \( E[X] = \lambda \) and \( \text{Var}[X] = \lambda \).
3. **Geometric** \( p \): \( p_X(k) = q^{k-1} p \) and \( E[X] = 1/p \) and \( \text{Var}[X] = q/p^2 \).
4. **Negative binomial** \( (n,p) \): \( p_X(k) = \binom{k-1}{n-1} p^n q^{k-n} \), \( E[X] = n/p \), \( \text{Var}[X] = nq/p^2 \).

**BASIC CONTINUOUS RANDOM VARIABLES**

1. **Uniform on** \([a,b] \): \( f_X(x) = 1/(b-a) \) on \([a,b]\) and \( E[X] = (a+b)/2 \) and \( \text{Var}[X] = (b-a)^2/12 \).
2. **Normal** \((\mu,\sigma^2)\): \( f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \) and \( E[X] = \mu \) and \( \text{Var}[X] = \sigma^2 \).
3. **Exponential** \( \lambda \): \( f_X(x) = \lambda e^{-\lambda x} \) (on \([0,\infty)\)) and \( E[X] = 1/\lambda \) and \( \text{Var}[X] = 1/\lambda^2 \).
4. **Gamma** \((n,\lambda)\): \( f_X(x) = \frac{\lambda^n}{\Gamma(n)} e^{-\lambda x} x^{n-1} \) (on \([0,\infty)\)) and \( E[X] = n/\lambda \) and \( \text{Var}[X] = n/\lambda^2 \).
5. **Cauchy**: \( f_X(x) = \frac{1}{\pi(1+x^2)} \) and both \( E[X] \) and \( \text{Var}[X] \) are undefined.
6. **Beta** \((a,b)\): \( f_X(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} \) on \([0,1]\) and \( E[X] = a/(a+b) \).

**MOMENT GENERATING / CHARACTERISTIC FUNCTIONS**

1. **Discrete**: \( M_X(t) = E[e^{itX}] = \sum_x p_X(x) e^{itx} \) and \( \phi_X(t) = E[e^{itX}] = \sum_x p_X(x) e^{itx} \).
2. **Continuous**: \( M_X(t) = E[e^{itX}] = \int_{-\infty}^{\infty} f_X(x) e^{itx} dx \) and \( \phi_X(t) = E[e^{itX}] = \int_{-\infty}^{\infty} f_X(x) e^{itx} dx \).
3. **If** \( X \) and \( Y \) **are independent**: \( M_{X+Y}(t) = M_X(t) M_Y(t) \) and \( \phi_{X+Y}(t) = \phi_X(t) \phi_Y(t) \).
4. **Affine transformations**: \( M_{aX+b}(t) = e^{bt} M_X(at) \) and \( \phi_{aX+b}(t) = e^{ibt} \phi_X(at) \).
5. **Some special cases**; if \( X \) is normal \((0,1)\), complete-the-square trick gives \( M_X(t) = e^{t^2/2} \) and \( \phi_X(t) = e^{-t^2/2} \). If \( X \) is Poisson \( \lambda \) get “double exponential” \( M_X(t) = e^{\lambda(e^{it}-1)} \) and \( \phi_X(t) = e^{\lambda(e^{it}-1)} \).

**WHY WE REMEMBER: BASIC DISCRETE RANDOM VARIABLES**

1. **Binomial** \((n,p)\): sequence of \( n \) coins, each heads with probability \( p \), have \( \binom{n}{k} \) ways to choose a set of \( k \) to be heads; have \( p^k(1-p)^{n-k} \) chance for each choice. If \( n = 1 \) then \( X \in \{0,1\} \) so \( E[X] = E[X^2] = p \), and \( \text{Var}[X] = E[X^2] - E[X]^2 = p - p^2 = pq \). Use expectation/variance additivity (for independent coins) for general \( n \).
2. **Poisson** \( \lambda \): \( p_X(k) = e^{-\lambda} \lambda^k / k! \) times \( k \)th term in Taylor expansion of \( e^\lambda \). Take \( n \) very large and let \( Y \) be \# heads in \( n \) tosses of coin with \( p = \lambda/n \). Then \( E[Y] = np \lambda \) and \( \text{Var}[Y] = npq \approx np \lambda \). Law of \( Y \) tends to law of \( X \) as \( n \to \infty \), so not surprising that \( E[X] = \text{Var}[X] = \lambda \).
3. **Geometric** \( p \): Probability to have no heads in first \( k-1 \) tosses and heads in \( k \)th toss is \((1-p)^{k-1}p\). If you are repeatedly tossing coin forever, makes intuitive sense that if you have \((\text{in expectation}) \) \( p \) heads per toss, then you should need \((\text{in expectation}) \) \( 1/p \) tosses to get a heads. Variance formula requires calculation, but not surprising that \( \text{Var}(X) \approx 1/p^2 \) when \( p \) is small (when \( p \) is small \( X \) is kind like of exponential random variable with \( p = \lambda \)) and \( \text{Var}(X) \approx 0 \) when \( q \) is small.
4. **Negative binomial** \((n,p)\): If you want \( n \)th heads to be on the \( k \)th toss then you have to have \( n-1 \) heads during first \( k-1 \) tosses, and then a heads on the \( k \)th toss. Expectations and variance are \( n \) times those for geometric (since were’re summing \( n \) independent geometric random variables).
WHY WE REMEMBER: BASIC CONTINUUM RANDOM VARIABLES

1. **Uniform on** $[a,b]$: Total integral is one, so density is $1/(b-a)$ on $[a,b]$. $E[X]$ is midpoint $(a+b)/2$. When $a=0$ and $b=1$, we know $E[X^2] = \int_0^1 x^2 dx = 1/3$, so that $\text{Var}(X) = 1/3 - 1/4 = 1/12$. Stretching out random variable by $(b-a)$ multiplies variance by $(b-a)^2$.

2. **Normal** $(\mu,\sigma^2)$: when $\sigma = 1$ and $\mu = 0$ we have $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. The function $e^{-x^2/2}$ is (up to multiplicative constant) its own Fourier transform. The fact that $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$ came from a cool and hopefully memorable trick involving passing to two dimensions and using polar coordinates. Once one knows the $\sigma = 1, \mu = 0$ case, general case comes from stretching/squashing the distribution by a factor of $\sigma$ and then translating it by $\mu$.

3. **Exponential** $\lambda$: Suppose $\lambda = 1$. Then $f_X(x) = e^{-x}$ on $[0,\infty)$. Remember the integration by parts induction that proves $\int_0^\infty e^{-x}x^n = n!$. So $E[X] = 1! = 1$ and $E[X^2] = 2! = 2$ so that $\text{Var}(X) = 2 - 1 = 1$. We think of $\lambda$ as rate (“number of buses per time unit”) so replacing 1 by $\lambda$ multiplies wait time by $1/\lambda$, which leads to $E[X] = 1/\lambda$ and $\text{Var}(X) = 1/\lambda^2$.

4. **Gamma** $(n,\lambda)$: Again, focus on the $\lambda = 1$ case. Then $f_X(x)$ is just $e^{-x} x^{n-1}$ times the appropriate constant. Since $X$ represents time until $n$th bus, expectation and variance should be $n$ (by additivity of variance and expectation). If we switch to general $\lambda$, we stretch and squash $f_X$ (and adjust expectation and variance accordingly).

5. **Cauchy**: If you remember that $1/(1+x^2)$ is the derivative of arctangent, you can see why this corresponds to the spinning flashlight story and where the $1/\pi$ factor comes from. Asymptotic $1/x^2$ decay rate is why $\int_{-\infty}^\infty f_X(x) dx$ is finite but $\int_{-\infty}^\infty f_X(x) x dx$ and $\int_{-\infty}^\infty f_X(x) x^2 dx$ diverge.

6. **Beta** $(a,b)$: $f_X(x)$ is (up to a constant factor) the probability (as a function of $x$) that you see $a-1$ heads and $b-1$ tails when you toss $a+b-2$ $p$-coins with $p = x$. So makes sense that if Bayesian prior for $p$ is uniform then Bayesian posterior (after seeing $a-1$ heads and $b-1$ tails) should be proportional to this. The constant $B(a,b)$ is by definition what makes the total integral one. Expectation formula (which you computed on pset) suggests rough intuition: if you have uniform prior for fraction of people who like new restaurant, and then $(a-1)$ people say they do and $(b-1)$ say they don’t, your revised expectation for fraction who like restaurant is $a/(a+b)$ (You might have guessed $\frac{(a-1)}{(a-1)+(b-1)}$, but that is not correct — and you can see why it would be wrong if $a-1 = 0$ or $b-1 = 0$.)
