STAT 24400 Lecture 8
4.1 The Expected Value of a Random Variable
4.2 Variance and Standard Deviation

Yibi Huang
Department of Statistics
University of Chicago

The Expected Value of a Random Variable

Example: A Card Game (from L03)

Consider a card game that you draw ONE card from a well-shuffled deck of cards. You win

- \$1 if you draw a heart,
- \$5 if you draw an ace (including the ace of hearts),
- \$10 if you draw the king of spades and
- \$0 for any other card you draw.

The PMF of your reward X is

Outcome	x	$p(x)$
Heart (not ace)	1	12/52
Ace	5	4/52
King of spades	10	1/52
All else	0	35/52

$$\Rightarrow \; p(x) = \begin{cases} 35/52 & \text{if } x = 0 \\ 12/52 & \text{if } x = 1 \\ 4/52 & \text{if } x = 5 \\ 1/52 & \text{if } x = 10 \\ 0 & \text{for all other values of } x \end{cases}$$

Long-Run Average of a Random Variable

If one plays the card game 5200 times (where the cards are drawn with replacement), then in the 5200 games, he is expected to get

- \$10 about 100 times (why?)
- \$5 about 400 times
- \$1 about 1200 times
- \$0 about 3500 times

His average reward in the 5200 games is hence about

$$\frac{\$10 \times 100 + \$5 \times 400 + \$1 \times 1200 + \$0 \times 3500}{5200}$$
$$= \$10 \times \frac{100}{5200} + \$5 \times \frac{400}{5200} + \$1 \times \frac{1200}{5200} + \$0 \times \frac{3500}{5200}$$
$$= \$10 \cdot \frac{1}{52} + \$5 \cdot \frac{4}{52} + \$1 \cdot \frac{12}{52} + \$10 \cdot \frac{35}{52} = \sum_x xp(x) = \$\frac{42}{52} \approx \$0.81$$

The long run average reward in a game is

$$\sum_x x \cdot p(x),$$

called the **expected value**, denoted as $\mathrm{E}(X)$ or μ_X.

Definition: Expected Value of a Discrete R.V.

Let X be a discrete random variable with PMF $p(x)$. The **expected value** or the **expectation** or the **mean** of X, denoted by $\mathrm{E}[X]$, or μ_x is a weighted average of the possible values of X, where the weights are the probabilities of those values.

$$\mu_x = \mathrm{E}[X] = \sum_{\text{all values of } x} x \cdot p(x)$$

provided that provided that $\sum_x |x| p(x) < \infty$. If the sum diverges, the expectation is undefined.

Expected Value of the Geometric Distribution

Recall the Geometric PMF is

$$p(k) = (1-p)^{x-1}p \quad \text{for } x = 1, 2, 3, \ldots.$$

To evaluate its expected value

$$E(X) = \sum_x x \cdot p(x) = \sum_{x=1}^{\infty} x(1-p)^{x-1}p,$$

we'll start from the geometric series

$$\sum_{x=0}^{\infty} r^x = \frac{1}{1-r} \quad \text{if } |r| < 1.$$

Differentiate both sides of the identity above with respect to r, we get another identity $\displaystyle\sum_{x=1}^{\infty} xr^{x-1} = \frac{1}{(1-r)^2}$, for $|r| < 1$.

Applying the new identity with $r = 1 - p$, we get

$$E(X) = \underbrace{\sum_{x=1}^{\infty} x(1-p)^{x-1}}_{=1/(1-(1-p))^2} p = \frac{p}{p^2} = \frac{1}{p}.$$

Example: Expected Value of Binomial

Recall the Binomial PMF is

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad 0 \le x \le n.$$

The expected value of the Binomial distribution is

$$\begin{aligned} \mathrm{E}(X) = \sum\nolimits_{x=0}^{n} x \cdot p(x) &= \sum\nolimits_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum\nolimits_{x=1}^{n} \underbrace{x \binom{n}{x}}_{\text{see below}} p^x (1-p)^{n-x} \end{aligned}$$

Key step:

$$\begin{aligned} x\binom{n}{x} = x\frac{n!}{x!(n-x)!} &= \frac{n!}{(x-1)!(n-x)!} \\ &= \frac{n \cdot (n-1)!}{(x-1)!(n-x)!} = n\binom{n-1}{x-1}. \end{aligned}$$

$$
\begin{aligned}
\mathrm{E}(X) &= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} n \binom{n-1}{x-1} p^x (1-p)^{n-x} \quad \left(\text{since } x\binom{n}{x} = n\binom{n-1}{x-1}\right) \\
&= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \\
&= np \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k}}_{=(p+1-p)^{n-1}=1} \quad (\text{let } k = x-1) \\
&= np
\end{aligned}
$$

where the last step comes from the Binomial expansion

$$
(a+b)^N = \sum_{k=0}^{N} \binom{N}{k} a^k b^{N-k}
$$

with $a = p$, $b = 1-p$, and $N = n-1$.

Expected Value of Negative Binomial

Recall the Negative Binomial PMF is

$$p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad x = r, r+1, \ldots.$$

The expected value of the Negative Binomial is

$$\mathrm{E}(X) = \sum_{x=r}^{\infty} \underbrace{x \binom{x-1}{r-1}}_{\text{see below}} p^r (1-p)^{x-r}$$

Key step:

$$x\binom{x-1}{r-1} = \frac{x \cdot (x-1)!}{(x-r)!(r-1)!} = \frac{x!}{(x-r)!(r-1)!}$$
$$= r \cdot \frac{x!}{(x-r)!(r-1)!r} = r\binom{x}{r}.$$

$$
\begin{aligned}
\mathrm{E}(X) &= \sum_{x=r}^{\infty} x \binom{x-1}{r-1} p^r (1-p)^{x-r} \\
&= \sum_{x=r}^{\infty} r \binom{x}{r} p^r (1-p)^{x-r} \quad \left(\text{since } x\binom{x-1}{r-1} = r\binom{x}{r}\right) \\
&= \frac{r}{p} \sum_{x=r}^{\infty} \binom{x}{r} p^{r+1} (1-p)^{x-r} \\
&= \frac{r}{p} \underbrace{\sum_{y=r+1}^{\infty} \binom{y-1}{r+1-1} p^{r+1} (1-p)^{y-(r+1)}}_{=1,\text{ since it's the sum of PMF for NB}(r+1,p)} \quad (\text{let } y = x+1) \\
&= \frac{r}{p}.
\end{aligned}
$$

$$
\begin{aligned}
E(X) &= \sum_{x=r}^{\infty} x \binom{x-1}{r-1} p^r (1-p)^{x-r} \\
&= \sum_{x=r}^{\infty} r \binom{x}{r} p^r (1-p)^{x-r} \quad \left(\text{since } x\binom{x-1}{r-1} = r\binom{x}{r}\right) \\
&= \frac{r}{p} \sum_{x=r}^{\infty} \binom{x}{r} p^{r+1} (1-p)^{x-r} \\
&= \frac{r}{p} \underbrace{\sum_{y=r+1}^{\infty} \binom{y-1}{r+1-1} p^{r+1} (1-p)^{y-(r+1)}}_{=1, \text{ since it's the sum of PMF for NB}(r+1,p)} \quad (\text{let } y = x+1) \\
&= \frac{r}{p}.
\end{aligned}
$$

Intuition: As it takes $1/p$ trials on average to get the first success, it'll take $r \cdot (1/p)$ trials on average to get r successes.

An Example Where $\mathsf{E}(X)$ Is Infinite

Game: You toss a fair coin repeatedly.
The longer you can keep getting heads, the more I'll reward you.
Specifically, if you get n consecutive Heads and then a Tail, I'll pay you 2^n cents. The PMF of the reward X (in cents) you get is

$$\mathrm{P}(X = 2^n) = \frac{1}{2^{n+1}}, \quad n = 0, 1, 2, \ldots.$$

The expected value of X is

$$\mathsf{E}(X) = \sum_{n=0}^{\infty} 2^n \mathrm{P}(X = 2^n) = \sum_{n=0}^{\infty} 2^n \cdot \frac{1}{2^{n+1}} = \sum_{n=0}^{\infty} \frac{1}{2} = \infty.$$

Definition: Expected Value of a Continuous R.V.

Let X be a continuous random variable with PDF $f(x)$. The **expected value** or the **expectation** or the **mean** of X, denoted by $\mathsf{E}[X]$, or μ_x is defined to be

$$\mu_x = \mathsf{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\mathrm{d}x$$

provided that provided that $\int_{-\infty}^{\infty} |x| f(x)\mathrm{d}x < \infty$. If the integral diverges, the expectation is undefined.

Expected Value — Gamma Distribution

Recall the PDF for Gamma(α, λ) is

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad \text{for } x \geq 0.$$

The expected value is

$$\begin{aligned}
\mathrm{E}(X) &= \int_0^\infty x \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathrm{d}x \\
&= \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{y}{\lambda}\right)^\alpha e^{-y} \frac{1}{\lambda} \mathrm{d}y \quad \left(\text{let } y = \lambda x \Rightarrow \mathrm{d}x = \frac{1}{\lambda}\mathrm{d}y\right) \\
&= \frac{1}{\lambda \Gamma(\alpha)} \underbrace{\int_0^\infty y^\alpha e^{-y} \mathrm{d}y}_{=\Gamma(\alpha+1)} = \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} = \frac{\alpha\Gamma(\alpha)}{\lambda\Gamma(\alpha)} = \frac{\alpha}{\lambda}.
\end{aligned}$$

Recall that the Gamma function $\Gamma(t)$ is defined to be

$$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} \mathrm{d}y,$$

and it has the property $\Gamma(t+1) = t\Gamma(t)$.

Expected Value — Beta Distribution

Recall the PDF for BETA(α, β) is

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad \text{for } 0 \le x \le 1.$$

Its expected value is

$$\mathrm{E}(X) = \int_0^1 x \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \mathrm{d}x = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \underbrace{\int_0^1 x^{\alpha}(1-x)^{\beta-1}\mathrm{d}x}_{=\mathrm{Beta}(\alpha+1,\beta)}.$$

Recall the Beta function Beta(u, v) is defined to be

$$\mathrm{Beta}(u,v) = \int_0^1 x^{u-1}(1-x)^{v-1}\mathrm{d}x, \text{ and it's equal to } \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}.$$

The expected value is thus

$$\mathrm{E}(X) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cancel{\Gamma(\beta)}} \frac{\Gamma(\alpha+1)\cancel{\Gamma(\beta)}}{\Gamma(\alpha+\beta+1)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\alpha\Gamma(\alpha)}{(\alpha+\beta)\Gamma(\alpha+\beta)} = \frac{\alpha}{\alpha+\beta}.$$

Expected Value — Normal Distribution

Recall the PDF for $X \sim N(\mu, \sigma^2)$ is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty.$$

Its expected value is

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{\mu + \sigma z}{\sqrt{2\pi}} e^{-z^2/2} \mathrm{d}z \quad \left(\text{let } z = \frac{x-\mu}{\sigma} \Rightarrow \mathrm{d}x = \sigma \mathrm{d}z\right) \\
&= \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \mathrm{d}z}_{=1 \text{ since it's intergal of normal PMF}} + \sigma \underbrace{\int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-z^2/2} \mathrm{d}z}_{=0 \text{ as } ze^{-z^2/2} \text{ is an odd function}} \\
&= \mu
\end{aligned}$$

Digression — Odd Function

A function $g(x)$ is called an *odd function* if it satisfies

$$g(-x) = -g(x), \quad \text{for all } x.$$

A function $h(x)$ is called an *even function* if it satisfies

$$h(-x) = -h(x), \quad \text{for all } x.$$

For an odd function $g(x)$,

$$\int_{-\infty}^{0} g(x)\mathrm{d}x \overset{\text{let} x=-y}{=} \int_{0}^{\infty} g(-y)\mathrm{d}y = -\int_{0}^{\infty} g(y)\mathrm{d}y,$$

and hence

$$\begin{aligned}\int_{-\infty}^{\infty} g(x)\mathrm{d}x &= \int_{-\infty}^{0} g(x)\mathrm{d}x + \int_{0}^{\infty} g(x)\mathrm{d}x \\ &= -\int_{0}^{\infty} g(x)\mathrm{d}x + \int_{0}^{\infty} g(x)\mathrm{d}x = 0.\end{aligned}$$

provided that $\int_{-\infty}^{\infty} |g(x)|\mathrm{d}x < \infty$.

If the PDF is an Even function (Symmetric About 0) ...

If the PDF $f(x)$ of a random variable X is an **even** function,

$$f(-x) = f(x) \quad \text{for all } x,$$

then

- $g(x) = xf(x)$ is an odd function since $g(-x) = -xf(-x) = -xf(x) = -g(x)$
- so $\mathrm{E}(X) = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x = 0$ provided $\int_{-\infty}^{\infty} |x|f(x)\mathrm{d}x < \infty$.

e.g., the **double exponential** distribution with the PDF

$$f(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < \infty,$$

has an expected value of 0 as the PDF is even.

Cauchy Distribution Has No Expected Value

Recall the PDF for Cauchy Distribution is

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty.$$

One might think $\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x = 0$ since $f(x)$ is an even function. In fact, its expected value doesn't exist since

$$\int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)} \mathrm{d}x = \infty.$$

Expected Values of Functions of Random Variables

Expected Values of Functions of Random Variables

If X is a random variable with PMF $p_X(x)$ or PDF $f_X(x)$, and $Y = g(X)$, how to find the expected value of Y?

Expected Values of Functions of Random Variables

If X is a random variable with PMF $p_X(x)$ or PDF $f_X(x)$, and $Y = g(X)$, how to find the expected value of Y?

Method 1: Find the PMF $p_Y(y)$ or PDF $f_Y(y)$ for Y and then calculate the expected value as

$$\mathrm{E}(Y) = \begin{cases} \sum_y y p_Y(y) & \text{if } Y \text{ is discrete,} \\ \int_{-\infty}^{\infty} y f_Y(y) \mathrm{d}y & \text{if } Y \text{ is continuous.} \end{cases}$$

Expected Values of Functions of Random Variables

If X is a random variable with PMF $p_X(x)$ or PDF $f_X(x)$, and $Y = g(X)$, how to find the expected value of Y?

Method 1: Find the PMF $p_Y(y)$ or PDF $f_Y(y)$ for Y and then calculate the expected value as

$$\mathrm{E}(Y) = \begin{cases} \sum_y y p_Y(y) & \text{if } Y \text{ is discrete,} \\ \int_{-\infty}^{\infty} y f_Y(y) \mathrm{d}y & \text{if } Y \text{ is continuous.} \end{cases}$$

Method 2: One can calculate $\mathrm{E}(Y)$ directly using the PMF or PDF of X as

$$\mathrm{E}(Y) = \mathrm{E}(g(X)) = \begin{cases} \sum_x g(x) p_X(x) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x & \text{if } X \text{ is continuous.} \end{cases}$$

Expected Values of Functions of Random Variables

If X is a random variable with PMF $p_X(x)$ or PDF $f_X(x)$, and $Y = g(X)$, how to find the expected value of Y?

Method 1: Find the PMF $p_Y(y)$ or PDF $f_Y(y)$ for Y and then calculate the expected value as

$$\mathrm{E}(Y) = \begin{cases} \sum_y y p_Y(y) & \text{if } Y \text{ is discrete,} \\ \int_{-\infty}^{\infty} y f_Y(y) \mathrm{d}y & \text{if } Y \text{ is continuous.} \end{cases}$$

Method 2: One can calculate $\mathrm{E}(Y)$ directly using the PMF or PDF of X as

$$\mathrm{E}(Y) = \mathrm{E}(g(X)) = \begin{cases} \sum_x g(x) p_X(x) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x & \text{if } X \text{ is continuous.} \end{cases}$$

- Method 2 is easier since one doesn't have to find the distribution of $Y = g(X)$, which is sometimes not easy
- Proof of the equivalence of the two methods requires advanced theory of integration

- Proof of the equivalence of the two methods for the discrete case is given on p.122 of the textbook
- For the continuous case, we will only prove the case that $g()$ is differentiable & strictly increasing. Recall in L04, we showed the PDF of $Y = g(X)$ in this case is

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y).$$

So

$$\begin{aligned}
\mathrm{E}(Y) &= \int_{-\infty}^{\infty} y f_Y(y) \mathrm{d}y \\
&= \int_{-\infty}^{\infty} \underbrace{y}_{=g(x)} \cdot f_X(\underbrace{g^{-1}(y)}_{=x}) \cdot \underbrace{\frac{d}{dy} g^{-1}(y) \mathrm{d}y}_{=\mathrm{d}x} \\
&= \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \mathrm{d}x
\end{aligned}$$

where the last equality comes from a change of variables $y = g(x)$, which implies

$$x = g^{-1}(y), \quad \text{and} \quad \mathrm{d}x = \frac{d}{dy} g^{-1}(y) \mathrm{d}y.$$

Expected Value of $aX + b$

If X is a random variable (discrete or continuous), the expected value for its Linear transformation $Y = g(X) = aX + b$ is

$$\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b.$$

Proof. We prove it for discrete X with PMF $p(x)$.
The proof for the continuous case is similar.

$$\begin{aligned}
\mathrm{E}(aX + b) &= \sum_x (ax + b)p(x) \\
&= \sum_x (ax\,p(x) + bp(x)) \\
&= \sum_x axp(x) + \sum_x bp(x) \\
&= a \underbrace{\sum_x xp(x)}_{=\mathrm{E}(X)} + b \underbrace{\sum_x p(x)}_{=1} \\
&= a\,\mathrm{E}(X) + b
\end{aligned}$$

Variance & Standard Deviation (SD)

One measure of spread of a random variable (or its probability distribution) is the *variance*.

The **variance** of a random variable X, denoted as $\text{Var}(X)$ or σ_X^2 is defined as the average squared distance from the expected value $\mu_X = \text{E}(x)$.

$$\begin{aligned}\text{Var}(X) &= \sigma^2 = \text{"sigma squared"} \\ &= \text{E}\left[(X - \mu_X)^2\right] \\ &= \begin{cases} \sum_x (x-\mu_X)^2 p_X(x) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\mathrm{d}x & \text{if } X \text{ is continuous.} \end{cases}\end{aligned}$$

provided that the variance is $< \infty$.

Square root of the variance is the *standard deviation (SD).*

$$\text{SD}(X) = \sigma = \sqrt{\text{Var}(X)}$$

Variance — Normal Distribution

The variance for $X \sim N(\mu, \sigma^2)$ is

$$\begin{aligned}\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathrm{d}x \\ &= \sigma^2 \int_{-\infty}^{\infty} \frac{z^2}{\sqrt{2\pi}} e^{-z^2/2} \mathrm{d}z \quad (\text{let } z = \frac{x-\mu}{\sigma} \Rightarrow \mathrm{d}x = \sigma \mathrm{d}z)\end{aligned}$$

It remains to find $\int_{-\infty}^{\infty} z^2 e^{-z^2/2} \mathrm{d}z$. Using integration by part and observe $z^2 e^{-z^2/2} = z \cdot \frac{d}{\mathrm{d}z}(-e^{-z^2/2})$, we get

$$\begin{aligned}\int_{-\infty}^{\infty} z^2 e^{-z^2/2} \mathrm{d}z &= \left[-z e^{-z^2/2}\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (-e^{-z^2/2}) \mathrm{d}z \\ &= 0 - 0 + \int_{-\infty}^{\infty} e^{-z^2/2} \mathrm{d}z = \sqrt{2\pi}.\end{aligned}$$

Plugging the above back to $\mathrm{Var}(X)$, we get

$$\mathrm{Var}(X) = \frac{\sigma^2}{\sqrt{2\pi}} \sqrt{2\pi} = \sigma^2.$$

Variance of $aX + b$

For $Y = aX + b$, we have proved that
$E(Y) = E(aX + b) = a\mu + b$, where $\mu = E(X)$ and hence

$$[Y - E(Y)]^2 = [(aX + b) - E(aX + b)]^2 = [aX + b - (a\mu + b)]^2 = a^2(X - \mu)^2.$$

Variance of $aX + b$

For $Y = aX + b$, we have proved that
$E(Y) = E(aX + b) = a\mu + b$, where $\mu = E(X)$ and hence

$$[Y - E(Y)]^2 = [(aX+b) - E(aX+b)]^2 = [aX+b-(a\mu+b)]^2 = a^2(X-\mu)^2.$$

Taking expected value of the above we get

$$\begin{array}{ccc} E[Y - E(Y)]^2 & = & E[a^2(X-\mu)^2] \\ \| & & \| \\ \mathrm{Var}(Y) & & a^2\, E[(X-\mu)^2] \\ \| & & \| \\ \mathrm{Var}(aX+b) & & a^2\, \mathrm{Var}(X) \end{array}$$

This shows that

$$\mathrm{Var}(aX + b) = a^2\, \mathrm{Var}(X).$$

Moment and Central Moment

Given a random variable X with mean μ (discrete or continuous),

- its *kth moment* is defined to be $\mathrm{E}[X^k]$, and
- its *kth centeral moment* is defined to be $\mathrm{E}[(X-\mu)^k]$,

provided that $\mathrm{E}[|X|^k] < \infty$ and $\mathrm{E}[|X-\mu|^k] < \infty$.

Note that

- the 1st moment $\mathrm{E}(X)$ is the mean = expected value
- the 1st central moment $\mathrm{E}(X-\mu)$ is always 0
- the 2nd central moment $\mathrm{E}[(X-\mu)^2]$ is the variance.

Moments of the Gamma Distribution

Recall the PDF for Gamma(α, λ) is $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$, for $x \geq 0$. Its kth moment is

$$\begin{aligned}
\mathrm{E}(X^k) &= \int_0^\infty x^k \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathrm{d}x = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+k-1} e^{-\lambda x} \mathrm{d}x \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\lambda}\right)^{\alpha+k-1} e^{-y} \frac{1}{\lambda} \mathrm{d}y \quad (\text{let } y = \lambda x \Rightarrow \mathrm{d}x = \frac{1}{\lambda} \mathrm{d}y) \\
&= \frac{1}{\lambda^k \Gamma(\alpha)} \underbrace{\int_0^\infty y^{\alpha+k-1} e^{-y} \mathrm{d}y}_{=\Gamma(\alpha+k)} = \frac{\Gamma(\alpha+k)}{\lambda^k \Gamma(\alpha)}.
\end{aligned}$$

Using the property $\Gamma(t+1) = t\Gamma(t)$ of the Gamma function, we get

$$\mathrm{E}(X^k) = \frac{\Gamma(\alpha+k)}{\lambda^k \Gamma(\alpha)} = \begin{cases} \alpha/\lambda & \text{if } k = 1 \\ \alpha(\alpha+1)/\lambda^2 & \text{if } k = 2 \\ \alpha(\alpha+1)(\alpha+2)/\lambda^k & \text{if } k = 3 \\ \prod_{i=1}^k (\alpha+k-1)/\lambda^k & \text{in general.} \end{cases}$$

A Shortcut Formula for Calculating Variance

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \mathrm{E}(X^2) - \mu^2$$

Proof. We'll prove the discrete case. The continuous case is similar.

$$\begin{aligned}
\mathrm{E}[(X-\mu)^2] &= \sum\nolimits_x (x-\mu)^2\, p(x) \\
&= \\
&= \underbrace{\qquad\qquad}\qquad \underbrace{\qquad\qquad}\qquad \underbrace{\qquad\qquad} \\
&= \qquad\qquad\qquad =
\end{aligned}$$

A Shortcut Formula for Calculating Variance

$$\text{Var}(X) = \text{E}[(X-\mu)^2] = \text{E}(X^2) - \mu^2$$

Proof. We'll prove the discrete case. The continuous case is similar.

$$\begin{aligned}
\text{E}[(X-\mu)^2] &= \sum_x (x-\mu)^2\, p(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2)\, p(x) \\
&= \underbrace{\qquad\qquad}\qquad \underbrace{\qquad\qquad}\qquad \underbrace{\qquad\qquad} \\
&= \qquad\qquad\qquad =
\end{aligned}$$

A Shortcut Formula for Calculating Variance

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \mathrm{E}(X^2) - \mu^2$$

Proof. We'll prove the discrete case. The continuous case is similar.

$$\begin{aligned}
\mathrm{E}[(X-\mu)^2] &= \sum_x (x-\mu)^2\, p(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2)\, p(x) \\
&= \underbrace{\sum_x x^2\, p(x)}_{=} - 2\mu \underbrace{\sum_x x\, p(x)}_{=} + \mu^2 \underbrace{\sum_x p(x)}
\end{aligned}$$

A Shortcut Formula for Calculating Variance

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \mathrm{E}(X^2) - \mu^2$$

Proof. We'll prove the discrete case. The continuous case is similar.

$$\begin{aligned} \mathrm{E}[(X-\mu)^2] &= \sum_x (x-\mu)^2\, p(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2)\, p(x) \\ &= \underbrace{\sum_x x^2\, p(x)}_{=\mathrm{E}(X^2)} - 2\mu \underbrace{\sum_x x\, p(x)} + \mu^2 \underbrace{\sum_x p(x)} \\ &= \qquad\qquad\qquad = \end{aligned}$$

A Shortcut Formula for Calculating Variance

$$\text{Var}(X) = \text{E}[(X-\mu)^2] = \text{E}(X^2) - \mu^2$$

Proof. We'll prove the discrete case. The continuous case is similar.

$$\begin{aligned}
\text{E}[(X-\mu)^2] &= \sum_x (x-\mu)^2\, p(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2)\, p(x) \\
&= \underbrace{\sum_x x^2\, p(x)}_{=\text{E}(X^2)} - 2\mu \underbrace{\sum_x x\, p(x)}_{=\mu} + \mu^2 \underbrace{\sum_x p(x)} \\
&= \qquad\qquad\qquad =
\end{aligned}$$

A Shortcut Formula for Calculating Variance

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \mathrm{E}(X^2) - \mu^2$$

Proof. We'll prove the discrete case. The continuous case is similar.

$$\begin{aligned}
\mathrm{E}[(X-\mu)^2] &= \sum_x (x-\mu)^2\, p(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2)\, p(x) \\
&= \underbrace{\sum_x x^2\, p(x)}_{=\mathrm{E}(X^2)} - 2\mu \underbrace{\sum_x x\, p(x)}_{=\mu} + \mu^2 \underbrace{\sum_x p(x)}_{=1} \\
&= \qquad\qquad\qquad =
\end{aligned}$$

A Shortcut Formula for Calculating Variance

$$\text{Var}(X) = \text{E}[(X-\mu)^2] = \text{E}(X^2) - \mu^2$$

Proof. We'll prove the discrete case. The continuous case is similar.

$$\begin{aligned}
\text{E}[(X-\mu)^2] &= \sum\nolimits_x (x-\mu)^2\, p(x) \\
&= \sum\nolimits_x (x^2 - 2\mu x + \mu^2)\, p(x) \\
&= \underbrace{\sum\nolimits_x x^2\, p(x)}_{=\text{E}(X^2)} - 2\mu \underbrace{\sum\nolimits_x x\, p(x)}_{=\mu} + \mu^2 \underbrace{\sum\nolimits_x p(x)}_{=1} \\
&= \text{E}(X^2) - 2\mu^2 + \mu^2 =
\end{aligned}$$

A Shortcut Formula for Calculating Variance

$$\text{Var}(X) = \text{E}[(X-\mu)^2] = \text{E}(X^2) - \mu^2$$

Proof. We'll prove the discrete case. The continuous case is similar.

$$\begin{aligned}
\text{E}[(X-\mu)^2] &= \sum\nolimits_x (x-\mu)^2\, p(x) \\
&= \sum\nolimits_x (x^2 - 2\mu x + \mu^2)\, p(x) \\
&= \underbrace{\sum\nolimits_x x^2\, p(x)}_{=\text{E}(X^2)} - 2\mu \underbrace{\sum\nolimits_x x\, p(x)}_{=\mu} + \mu^2 \underbrace{\sum\nolimits_x p(x)}_{=1} \\
&= \text{E}(X^2) - 2\mu^2 + \mu^2 = \text{E}(X^2) - \mu^2
\end{aligned}$$

Variance — Gamma Distribution

To find the variance for the Gamma distribution, we've obtained $\mathsf{E}(X^2) = \alpha(\alpha+1)/\lambda^2$ earlier, and so

$$\mathsf{Var}(X) = \mathsf{E}(X^2) - (\mathsf{E}(X))^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 = \frac{\alpha}{\lambda^2}.$$

It takes more work to calculate $\mathsf{E}(X-\mu)^2 = \mathsf{E}(X-\alpha/\lambda)^2$.