Expectation

X is r.v. with range $\mathcal{X}$.

Discrete:

$$Eg(X) = \sum_{x \in \mathcal{X}} g(x) \underbrace{f_X(x)}_{\text{pmf}}$$

if $\sum_{x \in \mathcal{X}} |g(x)| f_X(x) < \infty$

that is, previous sum converges absolutely.
Otherwise, $Eg(X)$ is undefined.

Continuous:

or $\int_{\mathcal{X}}$

$$Eg(X) = \int_{-\infty}^{\infty} g(x) \underbrace{f_X(x)}_{\text{pdf}} dx$$

if integral converges absolutely,
that is, $\int_{-\infty}^{\infty} |g(x)| f_X(x) dx < \infty$.

otherwise, $Eg(X)$ is undefined.

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = ?$$

$$\lim_{n \to \infty} \sum_{i=1}^{n} \frac{(-1)^{i-1}}{i} = \ln 2$$

but does <u>not</u> converge absolutely.

$$\int_0^\infty \frac{\sin x}{x}\,dx = ?$$

$$\lim_{L \to \infty} \int_0^L \frac{\sin x}{x}\,dx = \frac{\pi}{2}$$

but does <u>not</u> converge absolutely.

(exists)

$Eg(X)$ is well-defined only when $E|g(X)| < \infty$.

Examples where expected values do not exist.

Example

pmf $f_X(x) = \dfrac{1}{x(x+1)}$, $x = 1, 2, 3, \ldots$

Check: Is this a pmf?

$f_X(x) \geq 0$ for all x (obvious).

$$\sum_{x \in \mathcal{X}} f_X(x) = \sum_{x=1}^{\infty} \underbrace{\frac{1}{x(x+1)}}_{= \frac{1}{x} - \frac{1}{x+1}} \quad \left(\begin{array}{c}\text{Telescoping} \\ \text{Sum}\end{array}\right)$$

$$= \lim_{K \to \infty} \sum_{x=1}^{K} \left(\frac{1}{x} - \frac{1}{x+1}\right)$$

$$= \lim_{K \to \infty} \left(1 - \frac{1}{K+1}\right) = 1.$$

Yes, pmf.

Does EX exist?

$$E|X| = \sum_{x \in \mathcal{X}} |x| f_X(x)$$

($g(x) = x$ in this example)

$$= \sum_{x=1}^{\infty} x \cdot \frac{1}{x(x+1)} = \sum_{x=1}^{\infty} \frac{1}{x+1} = \infty$$

(harmonic series)

No. EX does not exist.
(or sometimes we say $EX = \infty$.)

Define $Y = (-1)^X X = \begin{cases} X & \text{if } X \text{ even} \\ -X & \text{if } X \text{ odd} \end{cases}$

Does EY exist?

$Y = g(X)$ with $g(x) = (-1)^x x$ so that EY exists if $E|g(X)| < \infty$.

$$\sum_{x \in \mathcal{X}} |g(x)| f_X(x) = \sum_{x=1}^{\infty} x \cdot \frac{1}{x(x+1)} = \infty$$

(same as above)

No. EY does not exist.

Example: Cauchy distribution

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

(is essentially same as t-distn. with 1 df.)

Does EX exist?

Again $g(x) = x$.

$$E|X| = \int |x| f_X(x)\,dx = \int_{-\infty}^{\infty} |x| \frac{1}{\pi} \frac{1}{1+x^2}\,dx$$

Eyeball this to see it diverges
(Book has formal argument. $\frac{x}{1+x^2}$ has a closed form anti-derivative.)

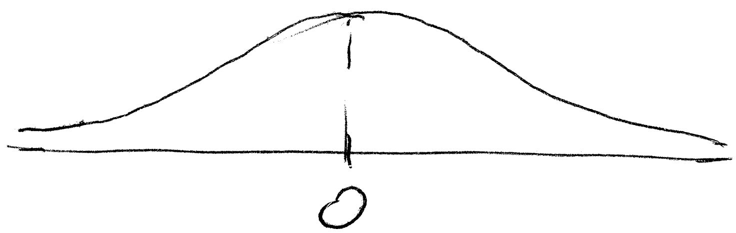
For large positive x, $\frac{x}{1+x^2} \simeq \frac{1}{x}$

and we know $\underbrace{\int_c^{\infty} \frac{1}{x}\,dx}_{= \log x \big|_c^{\infty}} = \infty$

So EX does not exist.

Cauchy example continued

Cauchy pdf is symmetric about zero.



Thus, $\int_{-K}^{K} x f_X(x)\,dx = 0$ for all K
(integrand is odd.)

Tempting to state that $EX = 0$.

But wrong!

Law of Large Numbers: If $X_1, X_2, \ldots, X_n$ is a large random sample from a population with mean $\mu = EX_i$, then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ will be very close to μ (usually).

For the Cauchy distn, the LLN fails because the Cauchy distn. does not have a mean.
(EX not defined.)

For Cauchy distn., $\bar{X} \overset{d}{=} X_1$!!

The Law of the Unconscious Statistician

Suppose $Y = g(X)$,
X and Y have pdf's, and
EY exists.

Then

$$EY = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx = E\,g(X).$$

That is, there are two ways to compute EY.

<u>Example</u>: Suppose X has pdf
$f_X(x) = 2x$ for $0 < x < 1$.

Then $E \log X = \int_0^1 (\log x)\, 2x\, dx = -\frac{1}{2}$.
(use integration by parts)

Alternatively,
$Y = \log X$ has range $\mathcal{Y} = (-\infty, 0)$
and pdf $f_Y(y) = f_X(e^y) \frac{d}{dy} e^y = 2e^y \cdot e^y$ for $y < 0$.

Thus $EY = \int_{-\infty}^{0} y \cdot 2e^{2y}\, dy = -\frac{1}{2}$.
(again, integrate by parts)

Important Special Cases of Expected Value

Expected Value	$Eg(X)$
Probabilities	$g(x) = 0$ or 1
Moments	$g(x) = x^k$
Moment generating functions (mgf's)	$g(x) = e^{tx}$

Probabilities as Expected Values

Notation:

Indicator functions are functions which take on only the values 0 or 1.

For $A \subset \mathbb{R}$ define the function

$$I_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{for } x \notin A. \end{cases}$$

Example

$$I_{(a,b]}(x) = \begin{cases} 1 & \text{for } a < x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

$I_A(\cdot)$ is a function.

Indicator random variables are random variables which take on only the values 0 or 1.

For $B \subset \Omega$ (B is an event) define the r.v.

$$I_B = \begin{cases} 1 & \text{if } B \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

I_B is a random variable.

Indicator random variables

Fact: For $B \subset \Omega$,

$$P(B) = E\, I_B .$$

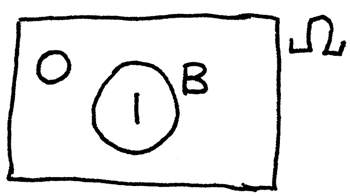
Proof: Let $Z = I_B$.

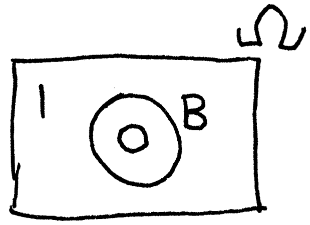
Z is discrete with pmf:

z	$f_Z(z)$
1	$P(B)$
0	$1-P(B)$

Thus
$$\begin{aligned} EZ &= \sum z f_Z(z) \\ &= 1 \cdot P(B) + 0 \cdot (1-P(B)) \\ &= P(B) \end{aligned}$$

Fact: $I_{B^c} = 1 - I_B$

Proof: I_B 

I_{B^c} 

Fact: $I_{A\cap B} = I_A \cdot I_B$ (and similarly for more events)

$$I_{A\cap B\cap C} = I_A \cdot I_B \cdot I_C$$

etc.

Proof: $I_A \cdot I_B = 1$

iff $I_A = 1$ and $I_B = 1$

iff A occurs and B occurs

iff $A\cap B$ occurs

iff $I_{A\cap B} = 1$

Properties of Indicator RV's

$$P(B) = E\, I_B$$

$$I_{B^c} = 1 - I_B$$

$I_{A\cap B} = I_A \cdot I_B$ (and similarly for more events)

Application: Show that

$$P(A \cap B^c \cap C^c) = P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C).$$

Argument:

$$P(A \cap B^c \cap C^c) = E(I_{A \cap B^c \cap C^c}).$$

$$\begin{aligned} I_{A \cap B^c \cap C^c} &= I_A \cdot I_{B^c} \cdot I_{C^c} \\ &= I_A \cdot (1 - I_B) \cdot (1 - I_C) \\ &= I_A - I_A I_B - I_A I_C + I_A I_B I_C \\ &= I_A - I_{A \cap B} - I_{A \cap C} + I_{A \cap B \cap C}. \end{aligned}$$

Now take expectations on both sides

$$\begin{aligned} P(A \cap B^c \cap C^c) &= E(I_A - I_{A \cap B} - I_{A \cap C} + I_{A \cap B \cap C}) \\ &= EI_A - EI_{A \cap B} - EI_{A \cap C} + EI_{A \cap B \cap C} \\ &= P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C). \end{aligned}$$

Moments (of a r.v. X)

$\mu'_k = E(X^k)$ = the k^{th} moment (about zero).

$\mu = \mu'_1 = EX$ = the mean of X.

$\mu_k = E[(X-\mu)^k]$ = the k^{th} central moment (the k^{th} moment about the mean).

$\sigma^2 = \mu_2 = E(X-\mu)^2$ = the variance of X = $\operatorname{Var} X$.

Existence of Moments

General definition says:

EX^k exists (is well defined) if $E|X|^k < \infty$.

<u>Note</u>: If X is a bounded r.v., then all moments exist.

<u>Proof</u>: Uses general fact that if $P(a \le g(X) \le b) = 1$, then $Eg(X)$ exists and

$$a \le Eg(X) \le b.$$

If X is bounded, then X^k is bounded for all $k = 1, 2, 3, \ldots$ so that EX^k exists.

In particular, if X is bounded between $-a$ and a, then X^k is bounded between $-a^k$ and a^k.

<u>Comment</u>: If X is <u>unbounded</u>, then EX^k may <u>or</u> may <u>not</u> exist. Work is required.

Uses of Moments

- **Descriptive Statistics**

  mean μ, variance σ^2

  (standardized) skewness $= \dfrac{\mu_3}{\sigma^3}$

  (standardized) kurtosis $= \dfrac{\mu_4}{\sigma^4} - 3$

  The skewness and kurtosis help describe the shape of a distribution.

- **Probability Inequalities (Bounds)**

  Markov's inequality

  If $P(X \geq 0) = 1$, then

  $$P(X \geq y) \leq \frac{EX}{y} \text{ for } y > 0.$$

  Chebyshev's inequality

  $$P(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2} \text{ for } t \geq 1.$$

- **Characterization of distributions**

  Example: If the moments of X agree with the moments of a normal distn, then X must have a normal distn.

Example: (Moments of a discrete distn.)

The Binomial Distribution

Suppose
  Coin with prob. π of heads.
  Toss it n times (tosses independent).
Define $X = \#$ of heads.

Then $X \sim \text{Binomial}(n, \pi)$ with pmf

$$f_X(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \quad x = 0, 1, \ldots, n.$$

Moments

EX^k is well-defined for $k = 1, 2, 3, \ldots$
since $\mathcal{X} = \{0, 1, \ldots, n\}$ is finite.

$$EX^k = \sum_{x \in \mathcal{X}} x^k f_X(x) = \sum_{x=0}^{n} x^k \binom{n}{x} \pi^x (1-\pi)^{n-x}.$$

Take $k=1$. Note that $x\binom{n}{x} = x \dfrac{n!}{x!(n-x)!}$

$$= \frac{n(n-1)!}{(x-1)!(n-x)!} = n\binom{n-1}{x-1}$$

for $x = 1, 2, \ldots, n$.

Thus $x\binom{n}{x} = \begin{cases} 0 & \text{for } x=0 \\ n\binom{n-1}{x-1} & \text{for } x=1,\ldots,n \end{cases}$

so that

$$EX = \sum_{x=0}^{n} x\binom{n}{x}\pi^{x}(1-\pi)^{n-x}$$

$$= 0 + \sum_{x=1}^{n} n\binom{n-1}{x-1}\pi^{x}(1-\pi)^{n-x}$$

$$= n\pi \sum_{x=1}^{n} \binom{n-1}{x-1}\pi^{x-1}(1-\pi)^{n-x}$$

(Note that $n-x = (n-1)-(x-1)$.
Terms in sum are similar to Binomial pmf.
Make the substitution $y = x-1$.

$$= n\pi \sum_{y=0}^{n-1} \underbrace{\binom{n-1}{y}\pi^{y}(1-\pi)^{n-1-y}}_{\text{Binomial}(n-1,\pi)\text{ pmf}}$$

$$= n\pi \cdot 1 = n\pi$$

Computation of EX^2

The text gives a direct calculation. Here is an alternative approach.

Note that

$$x(x-1)\binom{n}{x} = \begin{cases} 0 & \text{for } x = 0, 1 \\ n(n-1)\binom{n-2}{x-2} & \text{for } x = 2, \ldots, n \end{cases}$$

since

$$x(x-1)\binom{n}{x} = \frac{x(x-1)\, n!}{x!\,(n-x)!} = \frac{n(n-1)(n-2)!}{(x-2)!\,(n-x)!}$$

$$= n(n-1)\binom{n-2}{x-2} \qquad \text{for } 2 \le x \le n.$$

Thus

$$E\,X(X-1) = \sum_{x=0}^{n} x(x-1)\binom{n}{x}\pi^x(1-\pi)^{n-x}$$

$$= 0 + \sum_{x=2}^{n} n(n-1)\binom{n-2}{x-2}\pi^x(1-\pi)^{n-x}$$

$$= n(n-1)\pi^2 \sum_{x=2}^{n}\binom{n-2}{x-2}\pi^{x-2}(1-\pi)^{n-x}$$

Let $y = x-2$. Note $n-x = (n-2)-(x-2)$.

$$= n(n-1)\pi^2 \sum_{y=0}^{n-2}\underbrace{\binom{n-2}{y}\pi^y(1-\pi)^{n-2-y}}_{\text{Binomial}(n-2,\,\pi)\text{ pmf}}$$

$$= n(n-1)\pi^2 = E\,X(X-1).$$

Since $X^2 = X(X-1) + X$ we have

$$E X^2 = E\,X(X-1) + E X$$

$$= n(n-1)\pi^2 + n\pi .$$

Similarly, using

$$x(x-1)(x-2)\binom{n}{x} = n(n-1)(n-2)\binom{n-3}{x-3} \quad \text{for } x = 3, \ldots, n$$

$$= 0 \quad \text{for } x = 0, 1, 2$$

we can show

$$E\,X(X-1)(X-2) = n(n-1)(n-2)\,\pi^3$$

and use this to find $E X^3$.

Since

$$X^3 = X(X-1)(X-2) + 3\,X(X-1) + X$$

we have

$$E X^3 = E\,X(X-1)(X-2) + 3\,E\,X(X-1) + E X$$

$$= n(n-1)(n-2)\pi^3 + 3n(n-1)\pi^2 + n\pi$$

And so forth.

Moments for Continuous Distns.

The Gamma Distn.

$X \sim \text{Gamma}(\alpha, \beta)$ (with $\alpha > 0, \beta > 0$)

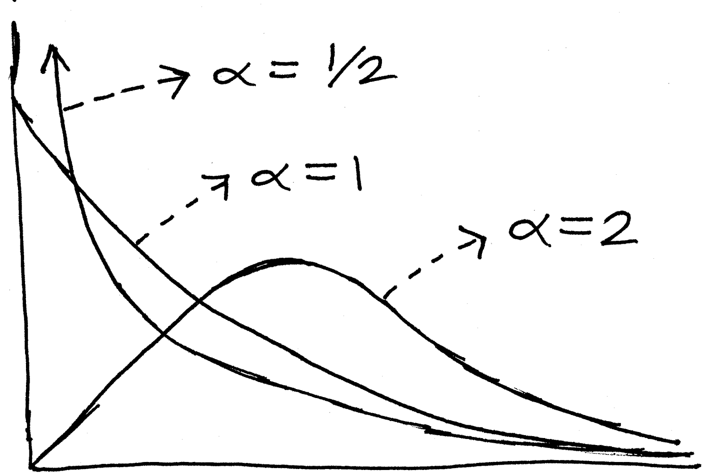
has pdf $f_X(x) = \dfrac{x^{\alpha-1} e^{-x/\beta}}{\underbrace{\beta^\alpha \Gamma(\alpha)}_{\text{normalizing constant}}}$ for $x > 0$.

normalizing constant
(makes integral one)

α is a "shape" parameter.
(Varying α changes the shape of the distn.)

β is a "scale" parameter.
(Varying β changes the scale of the distn.)

Special Cases



$(\alpha = 1, \beta = 1)$

$f_X(x) = e^{-x}$

$(\alpha = 1/2, \beta = 1)$

$f_X(x) = \dfrac{e^{-x}}{\sqrt{\pi}\sqrt{x}}$

$(\alpha = 2, \beta = 1)$

$f_X(x) = x e^{-x}$

Note:

For $0 < \alpha < 1$, $\lim_{x \downarrow 0} f_X(x) = \infty$.

For $\alpha = 1$, $\lim_{x \downarrow 0} f_X(x) = \frac{1}{\beta} > 0$.

For $\alpha > 1$, $\lim_{x \downarrow 0} f_X(x) = 0$.

The Gamma function

For $\alpha > 0$ define $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$.

Then

$$\Gamma(\alpha) = (\alpha - 1)! \quad \text{for } \alpha = 1, 2, 3, \ldots$$

$$\Gamma(\alpha + 1) = \alpha \, \Gamma(\alpha) \quad \text{for all } \alpha.$$

$$\Gamma(1/2) = \sqrt{\pi}$$

$$\Gamma(3/2) = \Gamma\left(\tfrac{1}{2} + 1\right) = \tfrac{1}{2} \Gamma\left(\tfrac{1}{2}\right) = \tfrac{1}{2}\sqrt{\pi}$$

$$\Gamma(5/2) = \Gamma\left(\tfrac{3}{2} + 1\right) = \tfrac{3}{2} \Gamma\left(\tfrac{3}{2}\right) = \tfrac{3}{2} \cdot \tfrac{1}{2} \sqrt{\pi}$$

etc.

Moments of the Gamma distn.

If $X \sim \text{Gamma}(\alpha, \beta)$, then

$$EX^K = \int_{-\infty}^{\infty} x^K f_X(x)\, dx$$

$$= \int_0^{\infty} x^K \cdot \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}\, dx$$

$$= \beta^K \frac{\Gamma(\alpha+K)}{\Gamma(\alpha)} \int_0^{\infty} \underbrace{\frac{x^{(\alpha+K)-1} e^{-x/\beta}}{\beta^{\alpha+K} \Gamma(\alpha+K)}}_{\text{pdf of Gamma}(\alpha+K,\, \beta)} dx$$

$$= \beta^K \frac{\Gamma(\alpha+K)}{\Gamma(\alpha)}$$

$$= \frac{\beta^K (\alpha+K-1)(\alpha+K-2) \cdots \alpha\, \Gamma(\alpha)}{\Gamma(\alpha)}$$

(by repeated use of $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$)

$$= \beta^K (\alpha+K-1)(\alpha+K-2) \cdots \alpha$$

Note: $\underset{(K=1)}{EX = \beta\alpha}$, $\underset{(K=2)}{EX^2 = \beta^2(\alpha+1)\alpha}$, etc.

For the Binomial and Gamma distns, all moments exist.

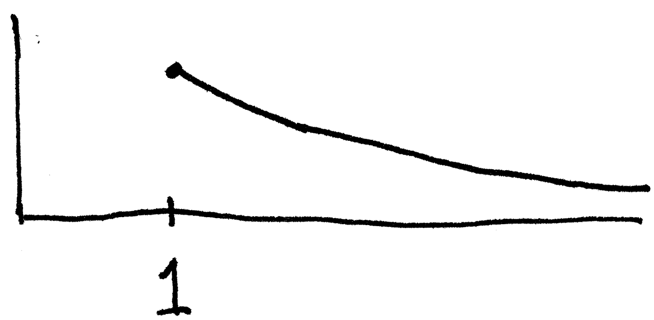
(EX^k finite for $k = 1, 2, 3, \ldots$)

For the Cauchy distn, EX does <u>not</u> exist and neither does EX^k for $k \geq 2$.

(<u>No</u> moments exist.)

<u>Example</u>: Special Case of Pareto distn.

pdf $f(x) = \dfrac{\beta}{x^{\beta+1}}$ for $x > 1$ $(\beta > 0)$.



Plot of $f(x)$

Suppose X has pdf $f(x)$.

(Then X is always positive so that $|X| = X$.)

EX^k exists if $E|X|^k < \infty$.

$$E|X|^k = \int_{-\infty}^{\infty} |x|^k f(x)\,dx$$

$$= \int_1^\infty x^k \frac{\beta}{x^{\beta+1}}\, dx$$

$$= \int_1^\infty \beta\, x^{k-\beta-1}\, dx$$

$$= \frac{\beta}{k-\beta}\, x^{k-\beta} \Big|_1^\infty \qquad \text{or } \beta \ln x \Big|_1^\infty \text{ if } k=\beta$$

$$= \frac{\beta}{\beta-k} \quad \text{if } k < \beta$$

$$= \infty \quad \text{if } k \geq \beta$$

Thus EX^k exists for $k<\beta$ and does not exist for $k \geq \beta$.

For example, if $\beta = 5$
EX^k exists only for $k = 1, 2, 3, 4$.

In General: (k_0, k are positive integers)

If EX^{k_0} does not exist, then EX^k for $k \geq k_0$ does not exist.

If EX^{k_0} does exist, then EX^k for $k \leq k_0$ does exist.