Week 5: Expected Values and Variances

Table of contents

(not be covered in exam/quizzes)

Probability Mass Function (PMF)

- The PMF $p(x)$ of a discrete random variable X is

$$p(x) = P(X = x)$$

- If $p(x)$ is the PMF of X, then

$$\sum_{\text{all x}} p(x) = 1.$$

Expected Value of a Discrete Random Variable

- The **expected value** (or **mean**, or **expectation**) of a discrete random variable X is

value probability

$$\boxed{E(X)} = \sum_{\text{all x}} xP(X = x) = \sum_{\text{all x}} xp(x)$$

- Notation

$$\mu_X = E(X), \text{ or } \mu = E(X) \text{ if the context is clear.}$$

Example 1

(a) Experiment: Roll a dice.

$X=$ number which shows up. What is $E(X)$?

$X = \{1, 2, 3, 4, 5, 6\}$

$\frac{1}{6} = P(X=1) = P(X=2) = P(X=3) = P(X=4) = P(X=5) = P(X=6)$

$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$

$= \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = \frac{7}{2} = 3.5$

Example 1

(b) Experiment: Throw a calibrated coin which has 1/3 chance of landing on head. X = number of heads which show up. What is $E(X)$?

$P(\text{head}) = \frac{1}{3}$ 1

$P(\text{tail}) = 1 - \frac{1}{3} = \frac{2}{3}$ 0

$X = \{0, 1\}$

value.

$P(X=1) = \frac{1}{3}$ and $P(X=0) = \frac{2}{3}$

prob.

$E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1)$

$= 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3}$

Example 2

A roulette wheel has the numbers 00,0,1,2,. . . ,36. 38

If you bet $1 that an odd number comes up, you win or lose $1 according to whether that event occurs. What is your expected net gain?

① odd-number : $\{1,3,5,7,\cdots 35\}$: $\frac{18}{38}$ $+1$

② non-odd-number : $\{00,0,2,4,\cdots 36\}$: $\frac{20}{38}$ -1

$X=\{-1,1\}$ $P(X=-1)=\frac{20}{38}$ $P(X=1)=\frac{18}{38}$

$E(X)=(-1)\left(\frac{20}{38}\right)+(1)\cdot\left(\frac{18}{38}\right)=-\frac{1}{19}$

Exercise 1: Bernoulli Distribution

Let $X \sim \text{Bernoulli}(p)$.

$X = \{0, 1\}$

(a) Write out the PMF $p(x)$ of X.

(b) Find $E(X)$.

X	0	1
$P(X=x)$	$1-P$	P

$E(X) = (0)(1-P) + (1)(P) = P$

Geometric distribution

Lemma 1

Let $X \sim \text{Geometric}(p)$, that is, $X = \#$ Bernoulli trials until the first success. Then

$$E(X) = \frac{1}{p}.$$

Proof. Optional.

Example 3

(a) If you toss a fair coin, how many tosses do you expect to have until you land on a head?

$P(\text{head}) = P(\text{tail}) = \frac{1}{2}.$

$X \sim \text{Geom}\left(\frac{1}{2}\right) \qquad p = \frac{1}{2}.$

$E(X) = \frac{1}{p} = \frac{1}{\frac{1}{2}} = 2.$

(b) If you roll a dice, how many rolls do you expect to have until you land on a six?

$X \sim \text{Geom}\left(\frac{1}{6}\right)$

$E(X) = \frac{1}{p} = \frac{1}{\frac{1}{6}} = 6$

Functions of random variables

Theorem 1

If $Y = g(X)$ is a function of a random variable X, then

$$E(Y) = \sum_{\text{all } x} \boxed{g(x)}\,\boxed{p(x)}.$$

probability

function

$$E[g(X)] = \sum_{\text{all } x} g(x)\,P(x)$$

Example 4

Let X be the number which shows up in a roll of a fair dice. Find $E(X^2)$.

$X=$	1	2	3	4	5	6
$X^2:$	1	4	9	16	25	36
$P=$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X^2) = 1\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 9\cdot\frac{1}{6} + 16\cdot\frac{1}{6} + 25\cdot\frac{1}{6} + 36\cdot\frac{1}{6}$$

$$= \frac{1}{6}\left[1+4+9+16+25+36\right] = \frac{91}{6}$$

not equal

$$\neq E(X)^2 = \left(\frac{7}{2}\right)^2 = \frac{49}{4}$$

Linearity of Expectations

Theorem 2

Let $X_1, \ldots, X_n$ be random variables and $a, a_1, \ldots, a_n$ be real numbers. Then

shift of expectation

$X_1 = \{1, 2, 3, 4, 5, 6\}$ $E(X) = 3.5 = \frac{7}{2}$

(a) $E(a + X_1) = a + E(X_1)$ $E(2+X_1)$

$2 + X_1 = \{3, 4, 5, 6, 7, 8\}$

(b) $E(a_1 X_1) = a_1 E(X_1)$ $E(2X_1) = 2E(X_1)$

(c) $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$

Linearity

$2X_1 : \{2, 4, 6, 8, 10, 12\}$

Properties of Expectation Function $E(X)$

- Linearity:

$$E(a + a_1X_1 + \cdots + a_nX_n) = a + a_1E(X_1) + \cdots + a_nE(X_n)$$

- Multiplicative (for $\boxed{\text{independent variables}}$): If X and Y are independent, then

$P(X \cap Y) = P(X)P(Y)$
both happens.

$$E(XY) = E(X)E(Y)$$

← only holds for independent X, Y.

Exercise 2

Use Theorem 2 to find the expected value of $X \sim \text{Binomial}(n, p)$.

n times in total.

$$E(X_1+X_2+X_3+\cdots+X_n) = E(X_1)+E(X_2)+E(X_3)+\cdots+E(X_n)$$
$$= p+p\cdots p = np$$

100 times of coin tosses.

$P(\text{head}) = P(\text{tail}) = \frac{1}{2}$.

$100 \times \frac{1}{2} = 50$.

Expected Value vs Variance

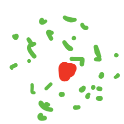
- The **expected value** of a random variable is
  - its **average value** and
  - is an indication of the **central value** of the PMF $p(x)$.

Expected Value vs Variance

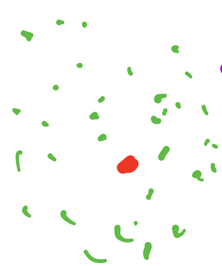
- The **expected value** of a random variable is
  - its **average value** and
  - is an indication of the **central value** of the PMF $p(x)$.
- The **variance** of a random variable is an indication of

  how dispersed the PMF $\mathbf{p(x)}$ **is about its center**.

lower variance ·



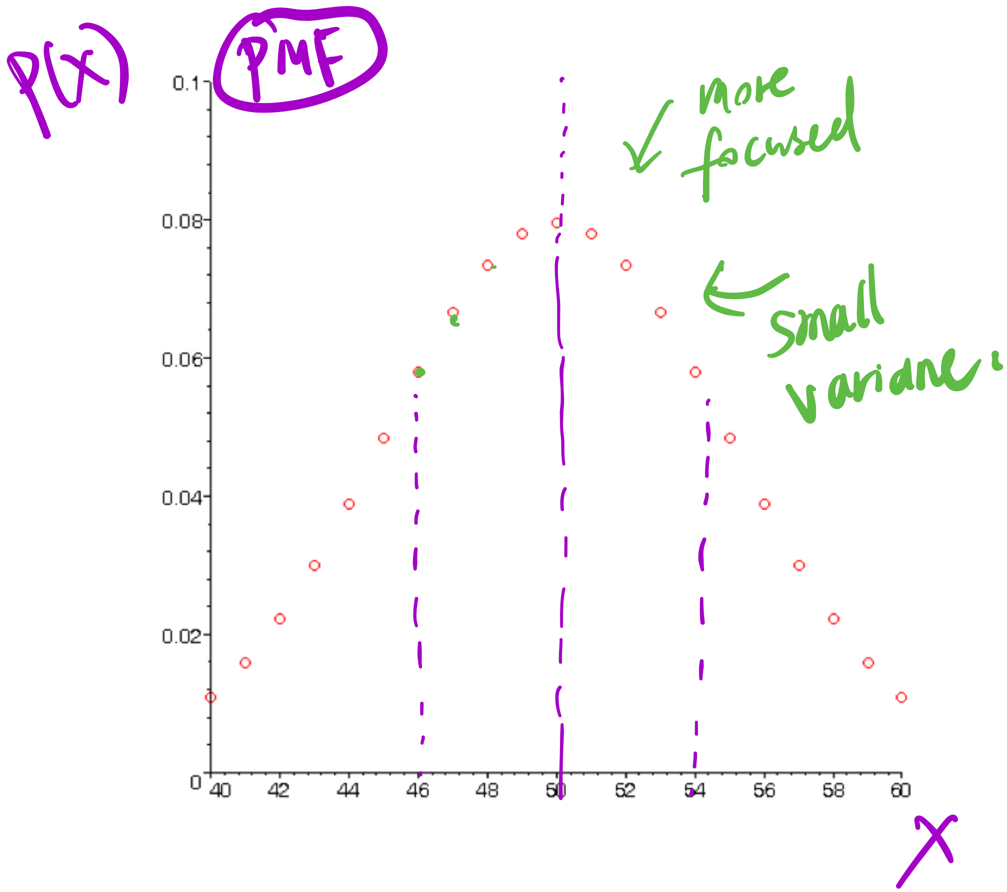
higher variance

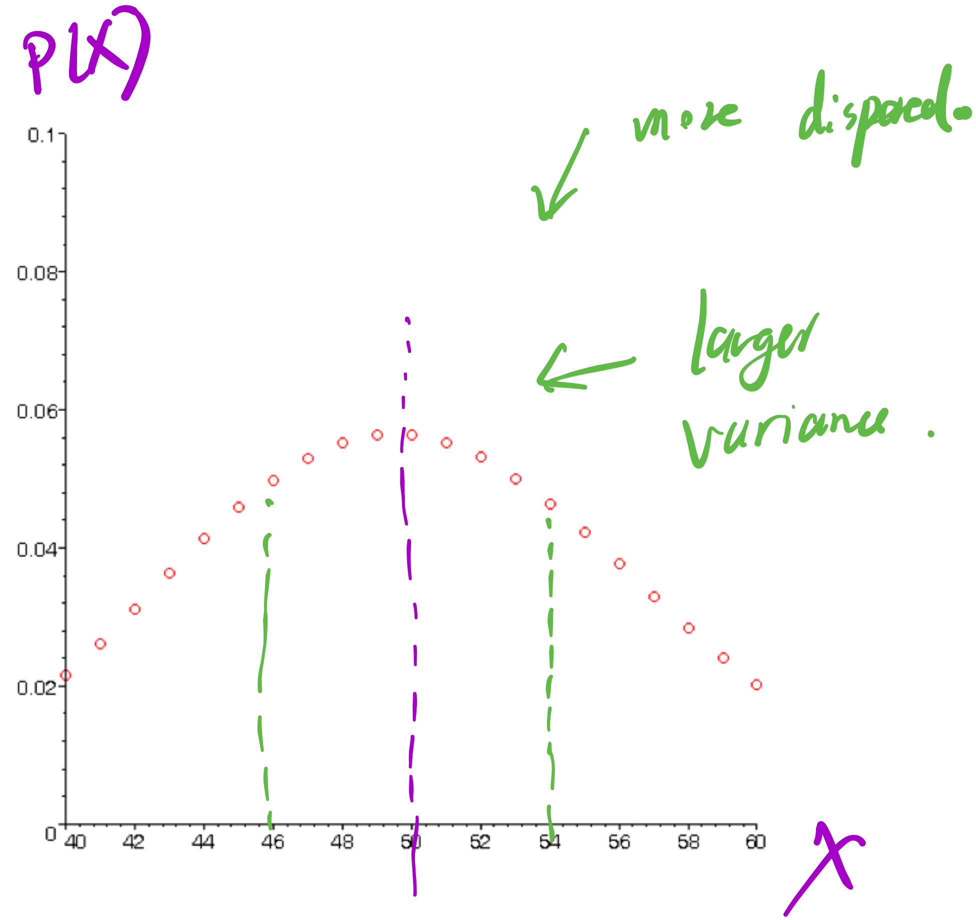


Example 5

What are the expected values in the following cases?

Which case has larger variance?





Variance and Standard Deviation

- If X is a variable with $\mu = E(X)$, the **variance** of X is

large if points all as dispersed.

$$\mathrm{Var}(X) = E[(\boxed{X - \mu})^2] = \sum_x (x-\mu)^2 p(x),$$

mean/center/average.

where the sum is taken over all possible values x of X.

Variance and Standard Deviation

- The **standard deviation** of X, denoted by σ_X, is

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

We simply write $\sigma = \sqrt{\mathrm{Var}(X)}$ if the context is clear.

Sigma

Example 6

$X = \#$ heads shown on a fair coin toss. Find $\text{Var}(X)$ and σ_X.

\# of heads.

$X = \quad 0 \qquad 1$

$P = \quad \frac{1}{2} \qquad \frac{1}{2}$

$E(X) = \frac{1}{2} = \mu$.

$$E[(X-\mu)^2] = (0-\tfrac{1}{2})^2 \cdot \tfrac{1}{2} + (1-\tfrac{1}{2})^2 \cdot \tfrac{1}{2}.$$

$$= \tfrac{1}{4} \cdot \tfrac{1}{2} + \tfrac{1}{4} \cdot \tfrac{1}{2} = \tfrac{1}{4} = \text{Var}(X)$$

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{\tfrac{1}{4}} = \tfrac{1}{2}.$$

Continue

Variance and Expectation

Theorem 3

The variance of X, if it exists, can be calculated as follows.

$$\underline{\mathrm{Var}(X)} = \boxed{\underline{E(X^2)}} - \boxed{\underline{[E(X)]}^2}. \rightarrow$$

derived from the definition.

Remark. In most cases, the formula

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

is used to compute $\mathrm{Var}(X)$.

Example 7

X = number shown on a toss of a fair dice. Find $\mathrm{Var}(X)$.

$\mathrm{Var}(X) = E(X^2) - (E(X))^2$.

$E(X) = \frac{7}{2} = 3.5$

$E(X^2) = \frac{91}{6}$.

$\mathrm{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$. variance.

$\sigma = \sqrt{\frac{35}{12}}$.

Example 8: Bernoulli(p)

Let $X \sim \text{Bernoulli}(p)$. Find $\text{Var}(X)$ and σ_X.

$X =$	0	1
$P =$	$1-p$	p
$X^2 =$	0^2	1^2

$E(X) = p$

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

$= p - p^2$

$= p(1-p)$

$\sigma_X = \sqrt{p(1-p)}$

Variance of $Y = a + bX$

Theorem 4

Let $a, b \in \mathbb{R}$. Then

ignore. just a shift.

$$\mathrm{Var}(a + \underline{bX}) = \mathrm{Var}(\underline{b}X) = b^2 \mathrm{Var}(X)$$

Proof. Optional.

Variance of Sum of Independent Variables

Theorem 5

If $X_1, \dots, X_n$ are $\boxed{\text{independent}}$ variables, then

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

not working for all

Proof. Optional.

Variables with Zero Variance

- The variance of X measures how it varies from its mean.
- If X is constant, then it does not vary at all. So $\mathrm{Var}(X) = 0$.
- Question: For what type of variables X do we have $\mathrm{Var}(X) = 0$?

Theorem 2. Let X be a random variable. Then

$$\mathrm{Var}(X) = 0 \Leftrightarrow \boxed{P(X = \mu) = 1} \text{ for some constant } \mu.$$

Means and variances of common distributions

Distribution	Mean	Variance
Bernoulli(p)	p	$p(1-p)$
Binomial(n,p)	np	$np(1-p)$
Geometric(p)	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson(λ)	λ	λ

given in quizzes and exams.

Remark. You can use all these results without proof.

Expected Values of a Continuous Random Variable

Let X be a continuous random variable with PDF $f(x)$. Then

PDF

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

$E(X) = \sum x\, p(x).$

Expectation of Uniform Distribution

The uniform distribution on the interval $[a, b]$ has PDF defined by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b, \\ 0 & \text{otherwise.} \end{cases}$$

Find $E(X)$.

$$E(X) = \int_a^b x \cdot f(x)\,dx = \int_a^b x \cdot \frac{1}{b-a}\,dx.$$

$$= \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\left(\frac{x^2}{2}\right)\Big|_a^b = \frac{1}{b-a}\left(\frac{b^2}{2} - \frac{a^2}{2}\right)$$

$$= \frac{1}{b-a}\left(\frac{b^2-a^2}{2}\right) = \frac{1}{\cancel{b-a}} \cdot \frac{(\cancel{b-a})(b+a)}{2} = \frac{a+b}{2}.$$

Expectation of Normal Distribution

standard deviation.

Let μ and σ be two real numbers with $\sigma > 0$. The variable $X \sim N(\mu, \sigma^2)$ has PDF

Variance.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ for any } x \in \mathbb{R}.$$

The expectation of normal distribution is $E(X) = \mu$.

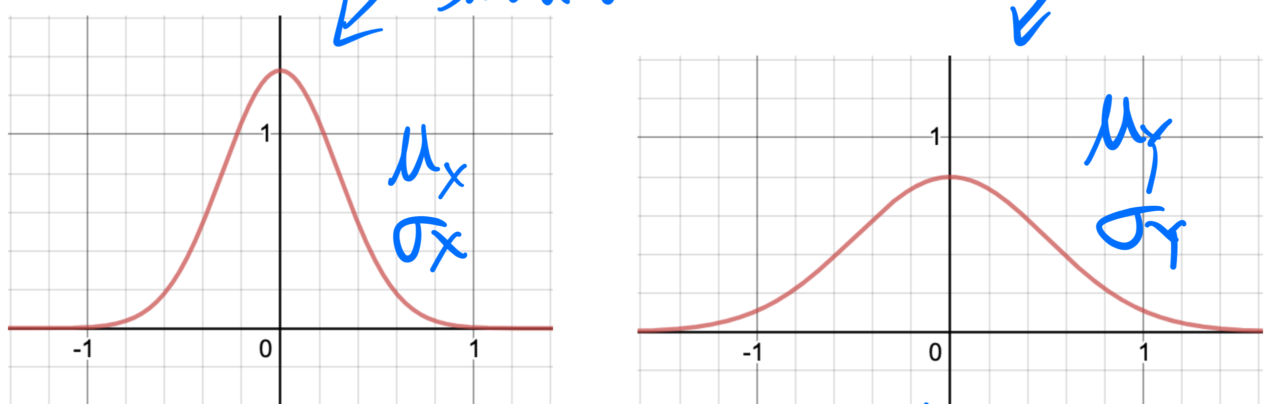
Expectation vs Variance

- The **expected value** of a random variable is its **average value** and can be viewed as an indication of the **central value** of the PDF or PMF.
- The **standard deviation** of a random variable is an indication of **how dispersed the probability distribution is about its center**.

Example 9

Consider the following graphs of two normal distributions both of those have mean value equal to 0.

Small variance. larger Variance.



$\mu_X = \mu_Y$ and $\sigma_X < \sigma_Y$

The RHS graph spreads wider than LHS graph $\Rightarrow$ the variance of the RHS distribution is larger than that of the LHS distribution.

Variance Formula

- If X is a discrete random variable, then

$Var(X)=E(X^2)-[E(X)]^2$

$$\mathrm{Var}(X) = \sum_x (x-\mu)^2 p(x),$$

where the sum is taken over all possible values x of X.

- If X is a continuous random variable, then

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx.$$

$$\int_{-\infty}^{+\infty} x^2 f(x)\,dx - \left(\int_{-\infty}^{+\infty} x f(x)\,dx\right)^2$$

Sum of independent variables

Theorem 3. If $X_1, \ldots, X_n$ are independent variables, then

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

Example 10

Determine the expected value and variance of the probability distribution over the specified range.

$$f(x) = \frac{1}{x^3}, \qquad 2 < x < 10$$

$E(X) = \int_2^{10} x \cdot f(x)\,dx = \int_2^{10} x \cdot \frac{1}{x^3}\,dx = \int_2^{10} \frac{1}{x^2}\,dx = \int_2^{10} x^{-2}\,dx.$

$= \left(\frac{x^{-2+1}}{-2+1}\right)\Big|_2^{10} = \left(\frac{x^{-1}}{-1}\right)\Big|_2^{10} = \left(-\frac{1}{x}\right)\Big|_2^{10}.$

$= -\frac{1}{10} - \left(-\frac{1}{2}\right) = \frac{2}{5}.$

Continue

$\mu = \frac{2}{5}$.

$$Var(X) = \sigma^2 = \int_2^{10} (x-\mu)^2 f(x) dx$$

$$\sigma^2 = \int_2^{10} \left(x - \frac{2}{5}\right)^2 \cdot \frac{1}{x^3} dx = \int_2^{10} \left(x^2 + \frac{4}{25} - \frac{4}{5}x\right) x^{-3} dx$$

$$= \int_2^{10} \left(x^{-1} + \frac{4}{25}x^{-3} - \frac{4}{5}x^{-2}\right) dx = \left(\ln|x| - \frac{2}{25}x^{-2} + \frac{4}{5}x^{-1}\right)\Big|_2^{10}$$

$$= \ln(5) - \frac{188}{625}$$

Exercise 2: $U(a,b)$

$\int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx$ or $\int_{-\infty}^{+\infty} x^2 f(x)\,dx - (\mu)^2$.

The uniform distribution on the interval $[a,b]$ has PDF defined by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b, \\ 0 & \text{otherwise.} \end{cases}$$

$\mu = \frac{a+b}{2}$.

Find $\mathrm{Var}(X)$ and σ_X.

Exercise 2 Solution

The mean μ of the uniform distribution is calculated as the expected value $E(X)$:

$$\mu = E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

$$= \frac{1}{b-a} \cdot \int_a^b x \, dx$$

$$= \frac{1}{b-a} \quad \frac{x^2}{2} = \frac{1}{b-a} \cdot \frac{b^2-a^2}{2} = \frac{a+b}{2}.$$

Exercise 2 Solution

The variance $\mathrm{Var}(X)$ of the uniform distribution is defined as

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

We need to find $E(X^2)$ first:

$$\frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{b-a}\cdot\frac{x^3}{3}\Big|_a^b = \frac{1}{b-a}\left(\frac{b^3-a^3}{3}\right).$$

$$= \frac{1}{\cancel{b-a}}\cdot\frac{\cancel{(b-a)}(b^2+ab+a^2)}{3} = \frac{b^2+ab+a^2}{3}$$

Exercise 2 Solution

$$\frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4}$$

$$= \frac{4b^2+4ab+4a^2-3a^2-6ab-3b^2}{12}$$

$$= \frac{b^2+a^2-2ab}{12} = \frac{(b-a)^2}{12}.$$

Variance of uniform distribution.

Exercise 3: Exp(λ)

The variable $X \sim \text{Exp}(\lambda)$ has PDF

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

(a) Prove that $E(X) = 1/\lambda$

(b) Prove that $\text{Var}(X) = 1/\lambda^2$.

Exercise 3 Solution

(a). $\int_0^{+\infty} x \cdot \lambda e^{-\lambda x}\, dx = \lambda \int_0^{+\infty} x e^{-\lambda x}\, dx$

IBP $\quad u = x \quad du = dx$

$dv = e^{-\lambda x}\, dx \quad v = -\frac{1}{\lambda} e^{-\lambda x}$

$= \lambda \lim_{t\to\infty} \int_0^t x e^{-\lambda x}\, dx$

$= \lambda \lim_{t\to\infty}\left[-\frac{x}{\lambda} e^{-\lambda x}\Big|_0^t - \int_0^t \left(-\frac{1}{\lambda}\right) e^{-\lambda x}\, dx\right]$

$= \lambda \lim_{t\to\infty}\left[\underbrace{\cancel{-\frac{t}{\lambda}e^{-\lambda t}}}_{\to 0} + 0 + \frac{1}{\lambda}\int_0^t e^{-\lambda x}\, dx\right]$

$= \lambda \lim_{t\to\infty}\left[\left(-\frac{1}{\lambda^2}\right) e^{-\lambda x}\Big|_0^t\right] = \lambda \lim_{t\to\infty}\left[-\frac{1}{\lambda^2}\underbrace{\cancel{e^{-\lambda t}}}_{0} + \frac{1}{\lambda^2}\right].$

$= \lambda \cdot \frac{1}{\lambda^2} = \frac{1}{\lambda}.$

Exercise 3 Solution

Means and Variances of Common Distributions

Distributions	Means	Variances
Bernoulli(p)	p	$p(1-p)$
Binomial(n,p)	np	$np(1-p)$
Geometric(p)	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson(λ)	λ	λ
U(a,b)	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$
N(μ,σ^2)	μ	σ^2
Exp(λ)	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

discrete

continuous

Function of Random Variables

Theorem 4.

Let X be a continuous random variable and let g be a function

$$E(g(X)) = \int g(x) f(x) dx$$

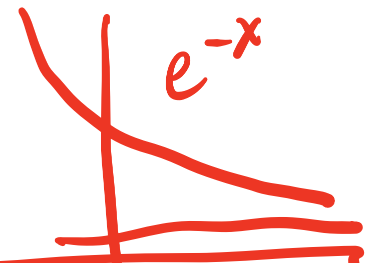
discrete : $E(g(x)) = \sum g(x) f(x)$

Example 11

Compute $E(X^2)$ for X that has PDF $f(x) = e^{-x}$, $x > 0$.

Using integration by parts

$$E(X^2) = \int_0^{\infty} \underbrace{x^2}_{g(x)} \underbrace{e^{-x}}_{f(x)} dx$$



$$\boxed{\lim_{x\to\infty} e^{-x} = 0}$$

IBP $\quad u = x^2 \quad du = 2x\,dx$

$dv = e^{-x}dx \quad v = -e^{-x}$

IBP again.

$u = x \quad du = dx$

$dv = e^{-x}dx \quad v = -e^{-x}$

$$\lim_{t\to\infty} \int_0^t x^2 e^{-x} dx = \lim_{t\to\infty} \left[-x^2 e^{-x} \Big|_0^t - \int_0^t -e^{-x} \cdot 2x\,dx \right]$$

$$= \lim_{t\to\infty} \left[-x^2 e^{-x} \Big|_0^t + 2\int_0^t x e^{-x} dx \right].$$

Example 11

Moment

A **moment** in statistics is a quantitative measure used to describe the shape of a probability distribution. Moments provide important characteristics of a distribution, such as its central tendency, spread, and shape. The n-th moment of a random variable X is essentially the expected value of X raised to the power n

Moment

- First moment (Mean) tells you where the center of the distribution is.
- Second moment (variance) tells you how spread out the distribution is.
- Third moment (skewness) tells you whether the distribution is symmetric or skewed
- Fourth moment (kurtosis) tells you how sharp or flat the distribution is compared to a normal distribution

Higher-order moments (third, fourth, etc) provide more detailed information about the shape of the distribution, but they are less commonly used in practice. They are important in risk management and finance.

Moment Generation Function

The **moment generating function (MGF)** of a random variable X is a function M_X in variable t defined by

$E'(e^{tx})$

$$M_X(t) = E(e^{tX}).$$

$= E(Xe^t) = E(X)$.

The moment generating function can be used to derive the moments of the distribution. The n-th moment of X is obtained by taking the n-th derivative of the MGF and evaluating at $\boxed{t = 0.}$

$E''(e^{tx})$

$= E'(Xe^{tx})$

$= E(X^2e^t)$ $E[e^{tx}]$

Example 12

1. The first derivative at $t = 0$ gives the mean: $\mu = E(X)$.
2. The second derivative at $t = 0$ gives the the second moment $E(X^2)$, from which the variance can be computed.

Concrete Formula

(a) The MGF of a discrete variable X with PMF $p(x)$ is

$$E(e^{tx}) = M_X(t) = \sum_x e^{tx} \boxed{p(x)}.$$ Probability

(b) The MGF of a continuous variable X with PDF $f(x)$ is

$$E(e^{tx}) = M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$