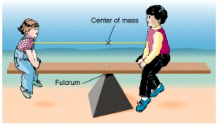
Chapter 4 sections

Summarizing distributions

- The distribution of X contains everything there is to know about the probabilistic properties of X.
- However, sometimes we want to summarize the distribution of X in one or a few numbers
  - e.g. to more easily compare two or more distributions.
- Examples of descriptive quantities:
  - Mean (= Expectation)
    - Center of mass - weighted average



  - Median, Moments
  - Variance, Interquartile Range (IQR), Covariance, Correlation

Definition of Expectation $\mu = E(X)$

Def: Mean aka. Expected value

Let X be a random variable with p(d)f $f(x)$. The *mean*, or *expected value* of X, denoted $E(X)$, is defined as follows

- X discrete:

$$E(X) = \sum_{\text{All } x} x f(x)$$

  assuming the sum exists.

- X continuous:

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

  assuming the integral exists.

If the sum or integral does not exists we say that the expected value does not exist.

The mean is often denoted with μ.

Examples

- Recall the distribution of $Y =$ the number of heads in 3 tosses (coin toss example from Lecture 4)

y	0	1	2	3
$f_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

then

$$E(Y) = 0\frac{1}{8} + 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = \frac{12}{8} = \frac{3}{2} = 1.5$$

Examples

- Recall the distribution of $Y =$ the number of heads in 3 tosses (coin toss example from Lecture 4)

y	0	1	2	3
$f_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

  then
$$E(Y) = 0\frac{1}{8} + 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = \frac{12}{8} = \frac{3}{2} = 1.5$$

- Find $E(X)$ where $X \sim \mathrm{Binom}(n, p)$. The pf of X is
$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \ldots, n$$

- Find $E(X)$ where $X \sim \mathrm{Uniform}(a, b)$. The pdf of X is
$$f(x) = \frac{1}{b-a} \quad \text{for } a \le x \le b$$

Expectation of $g(X)$

Theorem 4.1.1

Let X be a random variable with p(d)f $f(x)$ and $g(x)$ be a real-valued function. Then

- X discrete:
$$E\left(g(X)\right) = \sum_{\text{All } x} g(x)f(x)$$
- X continuous:
$$E\left(g(X)\right) = \int_{-\infty}^{\infty} g(x)f(x)\, dx$$

Example: Find $E(X^2)$ where $X \sim \mathrm{Uniform}(a, b)$.

Expectation of $g(X, Y)$

Theorem 4.1.2

Let X and Y be random variables with joint p(d)f $f(x, y)$ and let $g(x, y)$ be a real-valued function. Then

- X and Y discrete:

$$E\left(g(X,Y)\right) = \sum_{\text{All } x,y} g(x,y)f(x,y)$$

- X and Y continuous:

$$E\left(g(X,Y)\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy$$

Example: Find $E\left(\frac{X+Y}{2}\right)$ where X and Y are independent and $X \sim \mathrm{Uniform}(a, b)$ and $Y \sim \mathrm{Uniform}(c, d)$.

Properties of Expectation

Theorems 4.2.1, 4.2.4 and 4.2.6:

- $E(aX + b) = aE(X) + b$ for constants a and b.
- Let $X_1, \dots, X_n$ be n random variables, all with finite expectations $E(X_i)$, then

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

  - Corollary: $E\left(a_1X_1 + \cdots + a_nX_n + b\right) = a_1E(X_1) + \cdots + a_nE(X_n) + b$ for constants $b, a_1, \dots, a_n$.
- Let $X_1, \dots, X_n$ be n independent random variables, all with finite expectations $E(X_i)$, then

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$$

CAREFUL !!! In general $E(g(X)) \neq g(E(X))$.
For example: $E(X^2) \neq [E(X)]^2$

Examples

- If $X_1, X_2, \ldots, X_n$ are i.i.d. $\mathrm{Bernoulli}(p)$ random variables then $Y = \sum_{i=1}^n X_i \sim \mathrm{Binomial}(n, p)$.

$$E(X_i) = 0 \times (1-p) + 1 \times p = p \quad \text{for } i = 1, \ldots, n$$

$$\Rightarrow E(Y) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = np$$

Note: *i.i.d.* stands for *independent and identically distributed*

Definition of Variance $\sigma^2 = \mathrm{Var}(X)$

Def: Variance

Let X be a random variable (discrete or continuous) with a finite mean $\mu = E(X)$. The *Variance of X* is defined as

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

The *standard deviation of X is defined as* $\sqrt{\mathrm{Var}(X)}$

We often use σ^2 for variance and σ for standard deviation.

Theorem 4.3.1 – Another way of calculating variance

For any random variable X

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

Examples - calculating the variance

- Recall the distribution of $Y =$ the number of heads in 3 tosses (coin toss example from Lecture 4)

y	0	1	2	3
$f_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

We already found that $\mu = E(Y) = 1.5$. Then

$$\begin{aligned}\mathrm{Var}(Y) = (0-1.5)^2\frac{1}{8} + (1-1.5)^2\frac{3}{8} \\ + (2-1.5)^2\frac{3}{8} + (3-1.5)^2\frac{1}{8} \\ = 0.75\end{aligned}$$

- Find $\mathrm{Var}(X)$ where $X \sim \mathrm{Uniform}(a, b)$

Properties of the Variance

Theorems 4.3.2, 4.3.3, 4.3.4 and 4.3.5

- $\mathrm{Var}(X) \geq 0$ for any random variable X.
- $\mathrm{Var}(X) = 0$ if and only if X is a constant,
  i.e. $P(X = c) = 1$ for some constant c.
- $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$
- If $X_1, \ldots, X_n$ are independent we have

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

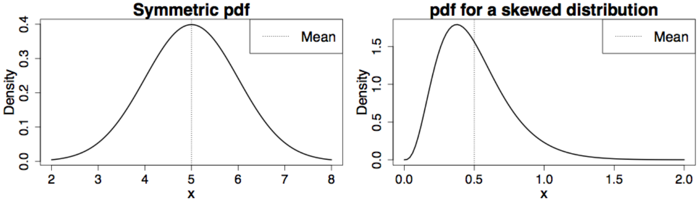
Examples

- If $X_1, X_2, \ldots, X_n$ are i.i.d. $\text{Bernoulli}(p)$ random variables then $Y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$.

$$\begin{aligned}
E(X_i) &= p \quad \text{for } i = 1, \ldots, n \\
E(X_i^2) &= 0^2 \times (1-p) + 1^2 \times p = p \quad \text{for } i = 1, \ldots, n \\
\Rightarrow \text{Var}(X_i) &= E(X_i^2) - [E(X_i)]^2 = p - p^2 = p(1-p) \\
\Rightarrow \text{Var}(Y) &= \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) = \sum_{i=1}^{n} p(1-p) \\
&= np(1-p)
\end{aligned}$$

Measures of location and scales

The mean is a measure of location, the variance is a measure of scale.



Moments and Central moments

Def: Moments

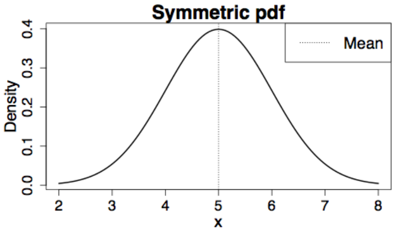
Let X be a random variable and k be a positive integer.

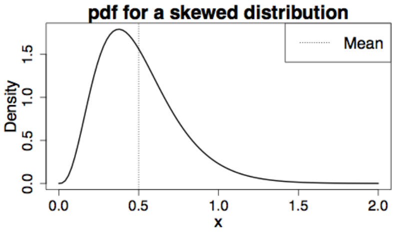
- The expectation $E(X^k)$ is called the *k^{th} moment of X*
- Let $E(X) = \mu$. The expectation $E\left((X-\mu)^k\right)$ is called the *k^{th} central moment of X*

- The first moment is the mean: $\mu = E(X^1)$
- The first central moment is zero: $E(X-\mu) = E(X) - E(X) = 0$
- The second central moment is the variance: $\sigma^2 = E\left((X-\mu)^2\right)$

Moments and Central moments

- *Symmetric distribution*: If the p(d)f $f(x)$ is symmetric with respect to a point x_0, i.e. $f(x_0 + \delta) = f(x_0 - \delta)$ for all δ
- It the mean of a symmetric distribution exists, then it is the point of symmetry.
- If the distribution of X is symmetric w.r.t. its mean μ then $E\left((X-\mu)^k\right) = 0$ for k odd (if the central moment exists)
- *Skewness*: $E\left((X-\mu)^3\right)/\sigma^3$





Moment generating function

Def: Moment Generating Function

Let X be a random variable. The function

$$\psi(t) = E\left(e^{tX}\right) \quad t \in \mathbb{R}$$

is called the *moment generating function (m.g.f.) of X*

Theorem 4.4.2

Let X be a random variables whose m.g.f. $\psi(t)$ is finite for t in an open interval around zero. Then the nth moment of X is finite, for $n = 1, 2, \ldots$, and

$$E(X^n) = \frac{d^n}{dt^n}\psi(t)\Big|_{t=0}$$

Example

Let $X \sim \mathrm{Gamma}(n, \beta)$. Then X has the pdf

$$f(x) = \frac{1}{(n-1)!\beta^n} x^{n-1} e^{-x/\beta} \quad \text{for } x > 0$$

Find the m.g.f. of X and use it to find the mean and the variance of X.

Properties of m.g.f.

Theorems 4.4.3 and 4.4.4:

- $\psi_{aX+b}(t) = e^{bt}\psi_X(at)$
- Let $Y = \sum_{i=1}^{n} X_i$ where $X_1, \ldots, X_n$ are independent random variables with m.g.f. $\psi_i(t)$ for $i = 1, \ldots, n$ Then

$$\psi_Y(t) = \prod_{i=1}^{n} \psi_i(t)$$

Theorem 4.4.5: Uniqueness of the m.g.f.

Let X and Y be two random variables with m.g.f.'s $\psi_X(t)$ and $\psi_Y(t)$.

If the m.g.f.'s are finite and $\psi_X(t) = \psi_Y(t)$ for all values of t in an open interval around zero, then X and Y have the same distribution.

Example

- Let $X \sim N(\mu, \sigma^2)$. X has the pdf

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

  and the m.g.f. for the normal distribution is

$$\psi(t) = \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$$

  Homework (not to turn in): Show that $\psi(t)$ is the m.g.f. of X.

- Let $X_1, \ldots, X_2$ be independent Gaussian random variables with means μ_i and variances σ_i^2.
  What is the distribution of $Y = \sum_{i=1}^n X_i$?