MATH 5010 –Foundations of Statistical Theory and Probability

❖ Expectations, Moments, and MGF

Instructor: He Wang
Department of Mathematics
Northeastern University

❖ Outline:

1. **Expectations**
2. **Moments**
3. **Moment Generating Functions**

❖ Expected Value

Expected value is a generalization of the concept "average".

The **expected value** or **mean** of a random variable X is defined to be

$$E[X] = \sum_{all\ k} k\, p_X(k)$$

Discrete random variable X

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)$$

Continuous random variable X

Example: Let X is the outcome of rolling a die.

k	$k=0$	$k=1$
$p_X(k)$	$1-p$	p

$$E[X] = \sum_{all\ k} k\, p_X(k) = 0(1-p) + 1(p) = p$$

The operational meaning is that $E[X]$ is the long-run average value of repeated measurements of the random variable X.

That is, suppose that we measure the random variable X in n independent trials, and record the results as $X_1, X_2, \ldots, X_n$. Then the long-run average value is

$$\overline{X_n} = \frac{1}{n}(X_1 + \cdots + X_n)$$

We will shortly see the Law of Large Numbers which implies that

$$\lim_{n \to \infty} \overline{X_n} = E(X)$$

Property: $E(aX + b) = aE(X) + b$

➢ Variance and Standard deviation

The **variance** of a random variable X is

$$Var(X) \coloneqq E\left[\left(X - E(X)\right)\right]^2$$

Variance is expected squared distance from the mean.

It measures the spread of the data.

The standard deviation is defined as the square root of the variance:

is $STD(X) \coloneqq \sqrt{Var(X)}$

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

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

❖ Expectations of a function

Recalled that the expected value of a discrete random variable X is

$$E[X] = \sum_{all\ k} k\, p_X(k)$$

The expected value of $Y = g(X)$ can be computed as

$$E[Y] = \sum_{all\ y} y\, p_Y(y) = \sum_{all\ y} yP(Y = y) = \sum_{all\ y} \sum_{\{k|y=g(x_k)\}} g(x_k)P(X = x_k)$$

$$= \sum_{all\ y} \sum_{k} \mathbb{I}\big(y = g(x_k)\big) g(x_k)P(X = x_k)$$

$$= \sum_{k} g(x_k)P(X = x_k) \sum_{y} \mathbb{I}\big(y = g(x_k)\big)$$

$$= \sum_{k} g(x_k)P(X = x_k)$$

So, for discrete random variable X

$$E[g(X)] = \sum_{k} g(x_k) P(X = x_k) = \sum_{x} g(x) p_X(x)$$

Similarly for continuous random variable X

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

Remark: $E[g(X)] \neq g(E[X])$

For example, $E[X^2] \neq (E[X])^2$ and $Var(X) := E[X^2] - (E[X])^2$

- Expectation are **linearly** decomposable, even if $X \text{ and } Y$ are dependent:

$$E[af(X) + bg(Y)] = aE[f(X)] + bE[g(Y)]$$

In particular,

$$E[aX + bY] = aE[X] + bE[Y]$$

- If $X \text{ and } Y$ are independent,

$$E[XY] = E[X]E[Y]$$

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$

❖ Moments

Definition: The $\boldsymbol{m}$**-th moment** of a random variable X is

$$E[X^m]$$

Definition: The $\boldsymbol{m}$**-th centered moment** of a random variable X is

$$E\left[\left(X - E(X)\right)^m\right]$$

Example: The **variance** $Var(X)$ is the second centered moment.

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

The square root of variance $\sigma \coloneqq \sqrt{Var(X)}$ is called the **standard deviation.**

Lists the means, variances some random variables.

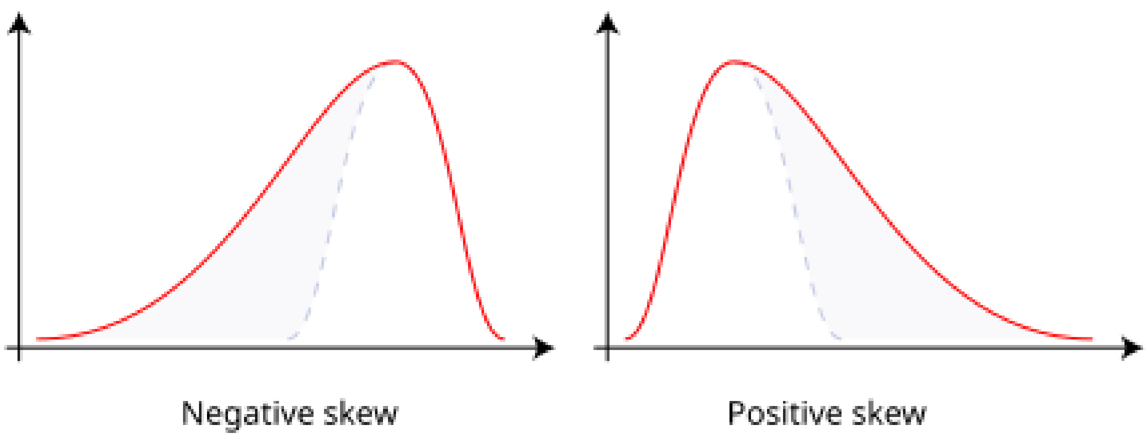
Name	Mean	Variance
$Bernoulli(p)$	p	$p(1-p)$
$Binomial(n, p)$	np	$np(1 - p)$
$Geometric(p)$	$np(1-p)$	$\frac{1-p}{p^2}$
$Poisson(\lambda)$	λ	λ
$Exponential(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$Gamma(n, \lambda)$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$
$Uniform(a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$Normal(\mu, \sigma^2)$	μ	σ^2
$Beta(\alpha, \beta)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

➢ Skewness

Skewness of a random variable is a measure of the asymmetry of the probability distribution.

Skewness can be calculated by the **third standard moments**

$$E\left[\left(\frac{X - E(X)}{\sigma}\right)^3\right] = \frac{E\left[\left(X - E(X)\right)^3\right]}{\sigma^3}$$

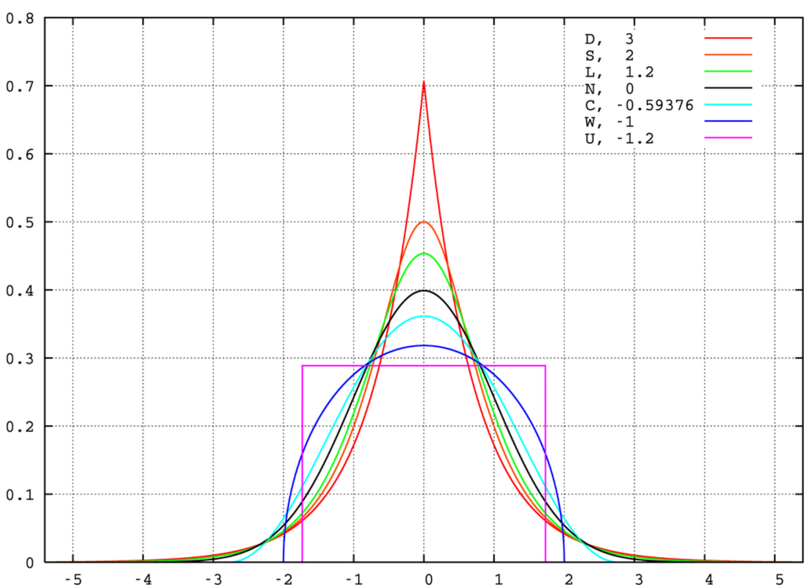


➢ Kurtosis

The **Kurtosis** is the fourth standardized moment, defined as

$$E\left[\left(\frac{X - E(X)}{\sigma}\right)^4\right] = \frac{E\left[\left(X - E(X)\right)^4\right]}{\sigma^4}$$

Kurtosis characterizes the “tailedness” of a distribution.



Laplace (D)ouble exponential distribution;
hyperbolic (S)ecant distribution;
(L)ogistic distribution;
(N)ormal distribution;
raised (C)osine distribution;
(W)igner semicircle distribution;
(U)niform distribution.

❖ Moment Generating Function (MGF)

Moment generating function (MGF) is a powerful function that describes the underlying features of a random variable.

Definition: The **MGF** of a random variable X is defined as

$$M_X(t) := E[e^{tX}]$$

When $M_X(t)$ exists in a neighborhood of 0, by Taylor expansion,

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots$$

$$M_X(t) = 1 + tE[X] + \frac{t^2 E[X^2]}{2!} + \frac{t^2 E[X^3]}{3!} + \cdots$$

So,

$$E[X^m] = M_X^{(m)}(0) = \left.\frac{d^m M_X(t)}{dt^m}\right|_{t=0}$$

All moments of X are generated by the function $M_X(t)$.

Theorem:

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

$M_{X+Y}(t) = M_X(t) M_Y(t)$ if $X\ and\ Y$ are independent.

Theorem: Suppose all moments exist for random variables $X\ and\ Y$.

1) If $X\ and\ Y$ have bounded support, the CDFs of $X\ and\ Y$ are equal if and only if all moments are equal.
2) If the moment generating functions exist and $M_X(t) = M_Y(t)$ for all t in an open neighborhood of 0, then the CDFs of $X\ and\ Y$ are equal.

Example (Bernoulli).

Let $X \sim Bernoulli(p)$. The MGF is

$$M_X(t) = E(e^{tX}) = pe^t + (1-p).$$

Example (Binomial).

Let $Y \sim Bin(n,p)$. The MGF is

$Y = X_1 + X_2 + \cdots + X_n$, with $X_i \sim Bernoulli(p)$ independent.

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = \left(pe^t + (1-p)\right)^n$$

Example (Poisson).

Let $X \sim Poisson(\lambda)$. Then its MGF is

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

Example (Exponential).

Let $X \sim Exp(\lambda)$. Then its MGF is

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

Example (Normal).

Let $X \sim N(\mu, \sigma^2)$.

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

Multivariate MGF

Consider a random vector $\vec{X} \in \mathbb{R}^d$

Definition: The **MGF** of a random vector $\vec{X} \in \mathbb{R}^d$ is defined as

$$M_X(\vec{t}) := E\left[e^{\vec{t}^T \vec{X}}\right]$$

Example: Multivariate Normal

Let $X \sim N(\vec{\mu}, \Sigma)$.

$$M_X(t) = E(e^{tX}) = e^{\vec{t}^T \vec{\mu} + \frac{1}{2} \vec{t}^T \Sigma \vec{t}}$$

So, a linear transformation $\vec{Z} = A\vec{X} + \vec{b} \sim N(b + A\vec{\mu}, \mathrm{A\Sigma A^T})$.

Example: Sum of Normal Distributions

Let $\vec{X} \sim N(\vec{\mu}_1, \Sigma_1)$.

$\vec{Y} \sim N(\vec{\mu}_2, \Sigma_2)$

The MGF of $\vec{Z} = \vec{X} + \vec{Y}$ is

$$M_{\vec{Z}}(\vec{t}) = E\left[e^{\vec{t}^T \vec{Z}}\right] = E\left[e^{\vec{t}^T \vec{X} + \vec{t}^T \vec{Y}}\right] = M_{(\vec{X}, \vec{Y})}(\vec{t}, \vec{t})$$

$$= \cdots$$

$$\Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{bmatrix}$$

$$= e^{\vec{t}^T(\vec{\mu}_1 + \overrightarrow{\mu_2}) + \frac{1}{2}\vec{t}^T(\Sigma_1 + \Sigma_2 + 2\Sigma_{12})\vec{t}}$$

$$\Sigma_{21} = COV(\vec{X}, \vec{Y})$$

References:

- **Book 1. [CB] Statistical Inference**, by Casella, George, Berger, Roger L, 2nd edition
- **Book 2. [W]: All of Statistics: Larry Wasserman**
- **Book 3. Introduction to Probability**. C.M. Grinstead and J.L. Snell. American Mathematical Society, 2012
- **Book 4. Introduction to Probability Models**, S. Ross, 12th edition (published by Academic Press).

https://online.stat.psu.edu/stat414/lesson/23/23.1