

Expectation and Moments

Yuki Shiraito

POLSCI 599 Statistical Methods in Political Research I
University of Michigan

Expectation

- Summary of r.v. X :
 - What is the gain you expect from a lottery?
 - What is the number of Dems you expect in a sample?
 - What is the lifetime income you expect from an academic job?
- **Expectation** or **expected value** of X : Weighted average of X where the weights are the probability measure of events $X = x$
- For a discrete r.v. X ,

$$\mathbb{E}[X] = \sum_x xf_X(x)$$

- For a continuous r.v. X ,

$$\mathbb{E}[X] = \int_x xf_X(x)dx$$

- $X \sim \text{Bern}(p) \Rightarrow \mathbb{E}[X] = p$
- $X \sim \text{Unif}(0, 1) \Rightarrow \mathbb{E}[X] = 1/2$

Existence of Expectation

- Expectation does not always exist
- **Existence of expectation:** $\mathbb{E}[X]$ exists if and only if $\mathbb{E}[X_-] < \infty$ or $\mathbb{E}[X_+] < \infty$, where $X_- \equiv -\min\{X, 0\}$ and $X_+ \equiv \max\{X, 0\}$.
 - 1 $\mathbb{E}[X_-] < \infty$ and $\mathbb{E}[X_+] < \infty$: $-\infty < \mathbb{E}[X] < \infty$
 - 2 $\mathbb{E}[X_-] < \infty$ and $\mathbb{E}[X_+] = \infty$: $\mathbb{E}[X] = -\infty$
 - 3 $\mathbb{E}[X_+] = \infty$ and $\mathbb{E}[X_-] < \infty$: $\mathbb{E}[X] = \infty$
 - 4 $\mathbb{E}[X_+] = \infty$ and $\mathbb{E}[X_-] = \infty$: $\mathbb{E}[X]$ does not exist
- Expectation can be infinity, but its sign should be well defined
- X follows the **standard Cauchy distribution:**

$$f_X(x) = \frac{1}{\pi(1+x^2)} \text{ for } -\infty < x < \infty$$

- Valid p.d.f: $\int_{-\infty}^{\infty} f(x)dx = [\tan^{-1}(x)/\pi]_{-\infty}^{\infty} = \{\pi/2 - (-\pi/2)\}/\pi = 1$
- $\int_0^{\infty} xf(x)dx = [\log(1+x^2)/2]_0^{\infty} = \infty$
- Similarly, $\int_{-\infty}^0 -xf(x)dx = [\log(1+x^2)/2]_{-\infty}^0 = \infty$
- Expectation does not exist for the Cauchy distribution

Indicator and Linearity

- **Expectation of Indicator:** For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let A be an event and define r.v. $1_A \equiv 1\{\omega \in A\}$. Then, $\mathbb{E}[1_A] = \mathbb{P}(A)$
- Corollary: Let $C \subset \mathbb{R}$. For r.v. X , define $1_C(X) \equiv 1\{X \in C\}$. Then, $\mathbb{E}[1_C(X)] = \mathbb{P}(X \in C)$
- Dice roll: X is the number on the face
 - Let $C \equiv \{2, 3, 4, 5\}$
 - $\mathbb{P}(X \in C) = 2/3$
 - $\mathbb{E}[1_C(X)] = 1 \times (4 \times 1/6) + 0 \times (2 \times 1/6) = 2/3$

- **Linearity:** Let X_1, X_2 be r.v.s. Then,

$$\mathbb{E}[aX_1 + bX_2 + c] = a\mathbb{E}[X_1] + b\mathbb{E}[X_2] + c$$

- Binomial expectation:

- By definition of expectation,

$$\mathbb{E}[X] = \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} = np \sum_{x=1}^n \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} = np$$

- Binom(n, p) is the distribution of $\sum_{i=1}^n X_i$ where $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Bern}(p)$

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = np$$

Random Vectors

- **Expectation of random vector:** For a random vector \mathbf{X} , its expectation is defined as

$$\mathbb{E}[\mathbf{X}] \equiv (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$$

where the expectation of X_i is over its marginal distribution

- Multinomial distribution $\mathbf{X} \sim \text{Multi}(n, \mathbf{p})$:
 - 1 Marginal distribution of X_1 is Binomial:

$$\begin{aligned} \mathbb{P}(X_1 = x_1) &= \sum_{x_2 \dots x_K} \frac{n!}{x_1! \dots x_K!} p_1^{x_1} \dots p_K^{x_K} \\ &= \frac{n!}{x_1!(n-x_1)!} p_1^{x_1} \sum_{x_2 \dots x_K} \frac{(n-x_1)!}{x_2! \dots x_K!} p_2^{x_2} \dots p_K^{x_K} \\ &= \frac{n!}{x_1!(n-x_1)!} p_1^{x_1} (1-p_1)^{n-x_1} \end{aligned}$$

- 2 $\mathbb{E}[\mathbf{X}] = (np_1, \dots, np_K)$

Functions and Product

- **Expectation of functions of r.v.:** Let X be a r.v. and $g : \mathbb{R} \rightarrow \mathbb{R}$. Then,

$$\mathbb{E}[g(X)] = \begin{cases} \sum_x g(x)f_X(x) & (X \text{ discrete}) \\ \int_x g(x)f_X(x)dx & (X \text{ continuous}) \end{cases}$$

Proof. Directly follows from the fact that for any $C \subset \mathbb{R}$,
 $\mathbb{P}(g(X) \in C) = \mathbb{P}(X \in \{x \in \mathbb{R} \mid g(x) \in C\})$

- X follows a **Geometric distribution**, $X \sim \text{Geom}(p)$:

$$f_X(x) = (1 - p)^{x-1}p, \text{ for } x = 1, 2, \dots$$

- ① St. Petersburg paradox: $g(x) \equiv 2^x \Rightarrow \mathbb{E}[g(X)] = \infty$ if $p = 1/2$
 - ② $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$ in general: $\mathbb{E}[X] = 2$
- Lemma: Let X be a discrete r.v. whose support is the non-negative integers. Then, $\mathbb{E}[X] = \sum_{x=1}^{\infty} \mathbb{P}(X \geq x)$
- **Product of independent r.v.s:** Let $X_i, i = 1, \dots, n$ are independent. Then, $\mathbb{E}[\prod_{i=1}^n X_n] = \prod_{i=1}^n \mathbb{E}[X_i]$
- $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Bern}(p) \Rightarrow \mathbb{P}(X_1 = 1, \dots, X_n = 1) = p^n$

Inequalities of Expectation

- If $X_1 \leq X_2$ with probability 1, i.e., $X_1(\omega) \leq X_2(\omega)$ for all $\omega \in \Omega$, then $\mathbb{E}[X_1] \leq \mathbb{E}[X_2]$
- If $a \leq X \leq b$ with probability 1, i.e., $a \leq X(\omega) \leq b$ for all $\omega \in \Omega$, then $a \leq \mathbb{E}[X] \leq b$
- **Jensen's inequality:** Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a concave (convex) function. Then, for a random vector X , $\mathbb{E}[g(X)] \leq (\geq)g(\mathbb{E}[X])$
- Concave function: A function $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is concave if and only if for every $a \in (0, 1)$,
$$g(ax + (1 - a)y) \geq ag(x) + (1 - a)g(y)$$
for any $x, y \in \mathbb{R}^n$
- Logarithm is common in statistics: $\mathbb{E}[\log(X)] < \log(\mathbb{E}[X])$

Moments and Variance

- **Moments of an r.v.:** For an r.v. X and a positive integer k , $\mathbb{E}[X^k]$ is called the k th *moment* of X
- Existence of moments: If $\mathbb{E}[X^k]$ exists, $\mathbb{E}[X^l]$ exists for any $l < k$
- **Central moments:** $\mathbb{E}[(X - \mathbb{E}[X])^k]$ is called the k th *central moment* or the k th *moment of X about the mean*
- If the k th moment exists, the l th central moment exists for $l \leq k$

- **Variance:** The second central moment of X is called the *variance* of X , denoted by $\mathbb{V}(X) \equiv \mathbb{E}[(X - \mathbb{E}[X])^2]$
- Variance and moments: $\mathbb{V}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
- $\mathbb{V}(X) \geq 0$, with equality if and only if $\mathbb{P}(X = c) = 1$ for some c
- $Y = aX + b \Rightarrow \mathbb{V}(Y) = a^2\mathbb{V}(X)$
- Variance of $\text{Bern}(p)$: $\mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p)$

Covariance

- **Covariance:** For r.v.s X_1 and X_2 , the *covariance* of X_1 and X_2 , denoted by $\text{Cov}(X_1, X_2)$, is defined as:

$$\text{Cov}(X_1, X_2) \equiv \mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])]$$

- Analogously to the variance, $\text{Cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$
- **Correlation:** The *correlation* of X_1 and X_2 , denoted by $\rho(X_1, X_2)$, is defined as:

$$\rho(X_1, X_2) \equiv \frac{\text{Cov}(X_1, X_2)}{\sqrt{\mathbb{V}(X_1)\mathbb{V}(X_2)}}$$

- $\rho(X_1, X_2) = \text{Cov} \left((X_1 - \mathbb{E}[X_1]) / \sqrt{\mathbb{V}(X_1)}, (X_2 - \mathbb{E}[X_2]) / \sqrt{\mathbb{V}(X_2)} \right)$
- Covariance depends on the scale of r.v.s, but $|\rho(X_1, X_2)| \leq 1$
- X_1 and X_2 are *uncorrelated* if and only if $\text{Cov}(X_1, X_2) = 0$
- X_1 and X_2 are independent $\Rightarrow X_1$ and X_2 are uncorrelated
- The converse does not necessarily hold:
 - $U \sim \text{Unif}(0, 1)$, $X_1 = \cos 2\pi U$ and $X_2 = \sin 2\pi U$
 - Clearly, X_1 and X_2 are not independent, but $\text{Cov}(X_1, X_2) = 0$
- Covariance and correlation indicate *linear* relationship b/w r.v.s

Variance-Covariance Matrix

- Trivially, $\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])] = \text{Cov}(X, X)$
- Variance-covariance Matrix:** For r.v.s X_1, \dots, X_n , we define the (variance-)covariance matrix, denoted by $\mathbb{V}(X)$ or Σ_X , as

$$\Sigma_X \equiv \begin{pmatrix} \mathbb{V}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \mathbb{V}(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \mathbb{V}(X_n) \end{pmatrix}$$

- In vector notation, $\Sigma_X = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]$
- Σ_X is positive semi-definite
- Σ_X is positive definite unless some r.v.s are constant \Rightarrow invertible
- If X_1, \dots, X_n are i.i.d., Σ_X is diagonal
- $\mathbb{V}(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 \mathbb{V}(X_i) + 2 \sum_{i < j} a_i a_j \text{Cov}(X_i, X_j)$
- X_i s are uncorrelated, $\mathbb{V}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbb{V}(X_i)$
- If X_1, \dots, X_n are i.i.d., $\mathbb{V}(\sum_{i=1}^n X_i) = \text{tr}(\Sigma_X)$

Conditional Expectation

- Summarizing prediction of Y using X :
 - Given $X = x$, Y is predicted by its conditional *distribution*
 - Conditional *expectation* is used to summarize the prediction
- Conditional expectation:** The *conditional expectation* of Y given X , denoted by $\mathbb{E}[Y | X]$, is the expectation of the conditional distribution of Y given X
- Conditional moments are defined with expectation replaced by conditional expectation, e.g. $\mathbb{V}(Y | X) \equiv \mathbb{E}[(Y - \mathbb{E}[Y | X])^2 | X]$
- Conditional expectation is an r.v.:*
 - $\mathbb{E}[Y | X]$ is a function of X
 - $\mathbb{E}[Y | X]$ has a distribution defined by F_X
- Conditional expectation is expectation:*
 - For any fixed x_1 , $\mathbb{E}[Y | X = x_1]$ is expectation
 - All the properties of expectation hold for $\mathbb{E}[Y | X = x_1]$
- Uniform-Binomial example:

$$\mathbb{E}[X_1 | X_2] = \int_0^1 x_1 \frac{x_1^{X_2} (1 - x_1)^{n - X_2}}{B(X_2 + 1, n - X_2 + 1)} dx_1 = \frac{X_2 + 1}{n + 2}$$

Properties of Conditional Expectation

- **Law of iterated expectations:** Let X and Y be r.v.s. Then,

$$\mathbb{E}[\mathbb{E}[g(X, Y) | X]] = \mathbb{E}[g(X, Y)]$$

for any function g .

- **Law of total variance:** Let X and Y be r.v.s. Then,

$$\mathbb{E}[\mathbb{V}(Y | X)] + \mathbb{V}(\mathbb{E}[Y | X]) = \mathbb{V}(Y)$$

- Uniform-Binomial example:

$$\mathbb{E}[X_2] = \mathbb{E}[nX_1] = \frac{n}{2}$$

$$\mathbb{V}[X_2] = \mathbb{E}[nX_1(1 - X_1)] + \mathbb{V}(nX_1) = \frac{n}{6} + \frac{n^2}{12}$$

- **Minimization of expected squared distance (regression):**

Conditional expectation is the “best” predictor in the sense that

$$\operatorname{argmin}_c \mathbb{E}[(Y - c)^2 | X] = \mathbb{E}[Y | X]$$

Proof.

$$\mathbb{E}[(Y - c)^2 | X] = \mathbb{E}[(Y - \mathbb{E}[Y | X])^2 | X] + \underbrace{(\mathbb{E}[Y | X] - c)^2}_{=0 \text{ iff } c = \mathbb{E}[Y | X]}$$

Standard Gaussian Distribution

- X follows the **standard multivariate Gaussian** distribution:

- Joint p.d.f.: For a vector of real numbers $x \equiv x_1, \dots, x_K$,

$$f_X(x) = \frac{1}{(2\pi)^{K/2}} e^{-\frac{1}{2}x^\top x} = \prod_{k=1}^K \frac{1}{\sqrt{2\pi}} e^{-\frac{x_k^2}{2}}$$

- Denoted by: $X \sim \mathcal{N}(0, \mathbf{I}_K)$
 - X_1, \dots, X_K are independent
 - $K = 1$: The standard Gaussian (Normal) distribution $\mathcal{N}(0, 1)$
- **Box-Muller Transformation**: Let U_1 and U_2 be independent uniform r.v.s. Define

$$X_1 = \sqrt{-2 \log U_1} \cos(2\pi U_2),$$

$$X_2 = \sqrt{-2 \log U_1} \sin(2\pi U_2).$$

Then,

$$X \sim \mathcal{N}(0, \mathbf{I}_2)$$

- Random number generator for the Gaussian distributions

Change of Variables

- **BH**, 8.1; **DS**, p. 172-3, 182-6
- **Change of variables**: Let X be a continuous random vector of length K and $Y \equiv g(X)$ where $g : \mathbb{R}^K \rightarrow \mathbb{R}^K$ is one-to-one and differentiable. Then, the p.d.f. of Y is

$$f_Y(y) = f_X(g^{-1}(y)) |\det(\mathbf{J}(y))|$$

where $g^{-1} : \mathbb{R}^K \rightarrow \mathbb{R}^K$ is the inverse function of g .

$\mathbf{J}(\cdot)$ is the Jacobian (matrix) of g^{-1} defined as:

$$\mathbf{J}(y) = \begin{pmatrix} \frac{\partial g_1^{-1}}{\partial y_1}(y) & \cdots & \frac{\partial g_1^{-1}}{\partial y_K}(y) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_K^{-1}}{\partial y_1}(y) & \cdots & \frac{\partial g_K^{-1}}{\partial y_K}(y) \end{pmatrix}$$

where $g_i^{-1}(y)$ is the i th element of $g^{-1}(y)$.

Univariate Change of Variables

- **Univariate change of variables:** Let X be a continuous r.v. and $Y \equiv g(X)$ where $g : \mathbb{R} \rightarrow \mathbb{R}$ is one-to-one and differentiable. Then, the p.d.f. of Y is

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}}{dy}(y) \right|$$

- **Proof.**

- 1 If $g(x)$ is one-to-one and differentiable, it is either strictly increasing or decreasing.
- 2 First, we assume that it is strictly increasing. Then, the c.d.f. of Y is

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

- 3 So the p.d.f. of Y is

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) \\ &= f_X(g^{-1}(y)) \frac{dg^{-1}}{dy}(y) \quad (\because \text{chain rule}) \end{aligned}$$

• **Proof, cont.**

- ④ Because g is strictly increasing, we have $\frac{dg^{-1}}{dy}(y) > 0$ so that

$$\frac{dg^{-1}}{dy}(y) = \left| \frac{dg^{-1}}{dy}(y) \right|.$$

- ⑤ Second, we consider the case in which g is strictly decreasing. Then, the c.d.f. of Y is

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) \\ &= 1 - \mathbb{P}(X \leq g^{-1}(y)) = 1 - F_X(g^{-1}(y)) \end{aligned}$$

Note that the inequality is flipped because g is strictly decreasing.

- ⑥ So the p.d.f. of Y is

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} (1 - F_X(g^{-1}(y))) \\ &= -f_X(g^{-1}(y)) \frac{dg^{-1}}{dy}(y) = f_X(g^{-1}(y)) \left(-\frac{dg^{-1}}{dy}(y) \right) \end{aligned}$$

- ⑦ Because g is strictly decreasing, we have $\frac{dg^{-1}}{dy}(y) < 0$ so that

$$-\frac{dg^{-1}}{dy}(y) = \left| \frac{dg^{-1}}{dy}(y) \right|.$$

Box-Muller Transformation

- Inverse of the Box-Muller transformation:

$$\begin{aligned} X_1 &= \sqrt{-2 \log U_1} \cos(2\pi U_2) & \Leftrightarrow & U_1 = e^{-(X_1^2 + X_2^2)/2} \\ X_2 &= \sqrt{-2 \log U_1} \sin(2\pi U_2) & & U_2 = \frac{1}{2\pi} \arctan\left(\frac{X_2}{X_1}\right) \end{aligned}$$

- The determinant of the Jacobian is:

$$\begin{aligned} \det(\mathbf{J}(\mathbf{X})) &= \det \begin{pmatrix} \frac{\partial U_1}{\partial X_1} & \frac{\partial U_1}{\partial X_2} \\ \frac{\partial U_2}{\partial X_1} & \frac{\partial U_2}{\partial X_2} \end{pmatrix} \\ &= \det \begin{pmatrix} -X_1 e^{-(X_1^2 + X_2^2)/2} & -X_2 e^{-(X_1^2 + X_2^2)/2} \\ \frac{-X_2/X_1^2}{2\pi(1+(X_2/X_1)^2)} & \frac{1/X_1}{2\pi(1+(X_2/X_1)^2)} \end{pmatrix} \\ &= \frac{-1 - X_2^2/X_1^2}{2\pi(1+(X_2/X_1)^2)} e^{-(X_1^2 + X_2^2)/2} \\ &= -\frac{1}{\sqrt{2\pi}} e^{-X_1^2/2} \times \frac{1}{\sqrt{2\pi}} e^{-X_2^2/2} \end{aligned}$$

Linear Transformation of Gaussian

- *Linear transformation*: Let X be a K -dimensional random vector. A linear transformation of X is

$$Y = a + \mathbf{A}X$$

where \mathbf{A} is a matrix with K columns

- **Multivariate Gaussian**: Let X follow the standard multivariate Gaussian distribution. Then, for a full rank $K \times K$ matrix \mathbf{A} and an K dimensional vector μ , $Y = \mu + \mathbf{A}X$ has a p.d.f.:

$$f_Y(y) = \frac{1}{(2\pi)^{K/2} \det(\Sigma)^{1/2}} e^{-\frac{1}{2}(y-\mu)^\top \Sigma^{-1}(y-\mu)}$$

where $\Sigma = \mathbf{A}\mathbf{A}^\top$

Proof.

$$X = \mathbf{A}^{-1}(Y - \mu)$$

$$J = \mathbf{A}^{-1}$$

- $\mathbb{E}[Y] = \mu$ and $\mathbb{V}(Y) = \Sigma$
- Uncorrelated \Leftrightarrow (pairwise) independent

Conditional and Marginal of Gaussian

- Let $(X_1, X_2)^\top \sim \mathcal{N}(\mu, \Sigma)$. The joint p.d.f. is:

$$\begin{aligned}
 f_{(X_1, X_2)}(x_1, x_2) &= \frac{e^{-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right\}}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\
 &= \underbrace{\frac{e^{-\frac{1}{2(1-\rho^2)\sigma_2^2} \left(x_2 - \mu_2 - \rho\sigma_2 \frac{x_1 - \mu_1}{\sigma_1} \right)^2}}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}}_{f_{X_2|X_1}(x_2|x_1)} \underbrace{\frac{e^{-\frac{1}{2} \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2}}{\sqrt{2\pi}\sigma_1}}_{f_{X_1}(x_1)}
 \end{aligned}$$

- $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), X_2 | X_1 \sim \mathcal{N}\left(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(X_1 - \mu_1), (1 - \rho^2)\sigma_2^2\right)$
- Both marginal and conditional distributions are Gaussian
- Regression $\mathbb{E}[X_2 | X_1] = \mu_2 + \frac{\rho\sigma_2}{\sigma_1}(X_1 - \mu_1)$ is linear in X_1
 \rightsquigarrow **linear regression**

Moment Generating Function

- Moment generating function:** Let X be an r.v. The *moment generating function (m.g.f.)* of X , denoted by $M_X(t)$, is defined as

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

if $\mathbb{E}[e^{tX}]$ exists for all $t \in (-s, s)$ for some $s > 0$.

- If m.g.f. is given, higher order moments can be easily computed:

$$\left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0} = \mathbb{E}[X^k e^{0X}] = \mathbb{E}[X^k], \text{ for } k = 1, 2, \dots$$

- If X_1, \dots, X_n are independent, the m.g.f. of $Y \equiv \sum_{i=1}^n X_i$ is

$$M_Y(t) = \mathbb{E}[e^{t \sum_{i=1}^n X_i}] = \mathbb{E}\left[\prod_{i=1}^n e^{tX_i}\right] = \prod_{i=1}^n \mathbb{E}[e^{tX_i}] = \prod_{i=1}^n M_{X_i}(t)$$

- M.g.f. uniquely determines the distribution:** If r.v.s X_1 and X_2 have m.g.f.s and $M_{X_1}(t) = M_{X_2}(t)$ for all $t \in (-a, a)$ for some $a > 0$, then c.d.f.s $F_{X_1}(x) = F_{X_2}(x)$ for all x

M.g.f. of Gamma Distributions

- Square of a Gaussian r.v. follows a *Gamma distribution*
- X follows a **Gamma distribution**:
 - P.d.f.:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad \text{for } x > 0$$

- Parameters: Shape $\alpha > 0$ and rate $\beta > 0$
 - Alternative parameterization: Shape $\alpha > 0$ and scale $\theta = 1/\beta$
 - Gamma function: $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$
 - Denoted by: $X \sim \text{Ga}(\alpha, \beta)$
- M.g.f. of the Gamma distribution:

$$\mathbb{E}[e^{tX}] = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta-t)x} dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta-t)^\alpha} = \left(\frac{\beta}{\beta-t} \right)^\alpha$$

- Expectation and variance: $\mathbb{E}[X] = \alpha/\beta$, $\mathbb{V}[X] = \alpha/\beta^2$
- Sum of independent Gamma r.v.s $X_i \sim \text{Ga}(\alpha_i, \beta)$, $i = 1, \dots, n$:

$$M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n \left(\frac{\beta}{\beta-t} \right)^{\alpha_i} = \left(\frac{\beta}{\beta-t} \right)^{\sum_{i=1}^n \alpha_i}$$

$$\Rightarrow \sum_{i=1}^n X_i \sim \text{Ga}(\sum_{i=1}^n \alpha_i, \beta)$$

Sample Moments

- Let X_1, \dots, X_n is a random sample from a distribution F_X
- In other words, X_1, \dots, X_n are *data*
- Sample moments:** Let X_1, \dots, X_n be i.i.d. r.v.s. The k th *sample moment*, denoted by M_k , is defined as

$$M_k \equiv \frac{1}{n} \sum_{i=1}^n X_i^k$$

- Why this is important: We never observe F_X , hence neither $\mathbb{E}[X^k]$
- We use M_k as an *estimator*—a function of r.v.s, therefore r.v.
- Mean and variance of M_k for any F_X :

$$\mathbb{E}[M_k] = \mathbb{E}[X^k], \quad \mathbb{V}(M_k) = \frac{\mathbb{V}(X^k)}{n}$$

- In particular, for **sample mean** $\bar{X} \equiv \sum_{i=1}^n X_i/n$,

$$\mathbb{E}[\bar{X}] = \mathbb{E}[X], \quad \mathbb{V}(\bar{X}) = \frac{\mathbb{V}(X)}{n}$$

- If we specify F_X , we can derive the distribution of M_k

Sample Mean of Gaussian R.v.

- We want to find the distribution of the sum of independent r.v.s
 \rightsquigarrow use m.g.f.!
- M.g.f. of the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$:

$$\begin{aligned}\mathbb{E}[e^{tX}] &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2+tx} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\{x-(\mu+\sigma^2t)\}^2+\mu t+\frac{1}{2}\sigma^2t^2} dx \\ &= e^{\mu t+\frac{1}{2}\sigma^2t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\{x-(\mu+\sigma^2t)\}^2} dx = e^{\mu t+\frac{1}{2}\sigma^2t^2}\end{aligned}$$

- M.g.f. of the sample mean:

$$\begin{aligned}\mathbb{E}[e^{t\bar{X}}] &= \prod_{i=1}^n \mathbb{E}[e^{\frac{t}{n}X_i}] = e^{n\left\{\mu\frac{t}{n}+\frac{1}{2}\sigma^2\left(\frac{t}{n}\right)^2\right\}} = e^{\mu t+\frac{1}{2}\left(\frac{\sigma^2}{n}\right)t^2} \\ \Rightarrow \bar{X} &\sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)\end{aligned}$$