S&DS 242/542: Theory of Statistics

Lecture 3: Probability review II

Moment generating functions

Moment generating functions

Last lecture we reviewed two ways of describing the **distribution** of a random variable:

- The *probability mass function* (PMF) or *probability density function* (PDF)
- The *cumulative distribution function* (CDF)

Today we discuss a third way to describe this distribution: the **moment generating function (MGF)**.

For any random variable X (discrete or continuous), its MGF $M_X(t)$ is defined for all $t \in \mathbb{R}$ by

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

Here $M_X(t) \in (0, \infty]$. Depending on the distribution of X, it is possible that $M_X(t) = \infty$ for some values of t.

Poisson MGF

Example: Let $X \sim \text{Poisson}(\lambda)$ be a Poisson random variable.

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

$$= \sum_{x=0}^{\infty} e^{tx} \cdot \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \underbrace{\sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}}_{\sum_{x=0}^{\infty} \frac{a^x}{x!} = e^a} = e^{-\lambda} \cdot e^{\lambda e^t} = \underbrace{e^{\lambda(e^t - 1)}}_{\text{MGF of Poisson}(\lambda)}$$

Standard normal MGF

Example: Let $X \sim \mathcal{N}(0,1)$ be a standard normal random variable.

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

$$= \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\underbrace{-\frac{x^2}{2} + tx}_{-\frac{1}{2}(x-t)^2 + \frac{t^2}{2}}}\, dx$$

$$= \int_{-\infty}^{\infty} e^{t^2/2} \cdot \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2}}_{\text{PDF of } \mathcal{N}(t,1)}\, dx = \underbrace{e^{t^2/2}}_{\text{MGF of } \mathcal{N}(0,1)}$$

General normal MGF

Example: Let $X \sim \mathcal{N}(\mu, \sigma^2)$ be any normal random variable. Then we may *represent* $X = \mu + \sigma Z$ where $Z \sim \mathcal{N}(0,1)$.

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

$$= \mathbb{E}[e^{t(\mu + \sigma Z)}]$$

$$= \mathbb{E}[e^{t\mu} \cdot e^{t\sigma Z}] = e^{t\mu} \underbrace{\mathbb{E}[e^{t\sigma Z}]}_{= M_Z(t\sigma) = e^{\frac{t^2\sigma^2}{2}}}$$

$$= \underbrace{e^{t\mu + \frac{t^2\sigma^2}{2}}}_{\text{MGF of } \mathcal{N}(\mu, \sigma^2)}$$

Gamma MGF

Example: Let $X \sim \text{Gamma}(\alpha, \beta)$ be a Gamma random variable.

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

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

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

$$= \begin{cases} \infty & \text{if } t \geq \beta \\ \frac{\Gamma(\alpha)}{(\beta-t)^\alpha} \int_0^\infty \underbrace{\frac{(\beta-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\beta-t)x}}_{\text{PDF of Gamma}(\alpha,\, \beta-t)} \, dx & \text{if } t < \beta \end{cases}$$

$$= \begin{cases} \infty & \text{if } t \geq \beta \\ \frac{1}{(1-\frac{t}{\beta})^\alpha} & \text{if } t < \beta \end{cases} \quad \leftarrow \text{MGF of Gamma}(\alpha, \beta).$$

The MGF encodes the moments of X

$M_X(t)$ is called the Moment Generating Function because its Taylor expansion encodes the moments of X:

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \mathbb{E}\Big[1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \dots\Big] \\ &= 1 + t \underbrace{\mathbb{E}[X]}_{1^{\text{st}}\text{ moment}} + \frac{t^2}{2!} \underbrace{\mathbb{E}[X^2]}_{2^{\text{nd}}\text{ moment}} + \frac{t^3}{3!} \underbrace{\mathbb{E}[X^3]}_{3^{\text{rd}}\text{ moment}} + \dots \end{aligned}$$

Computing its derivatives:

$$M_X'(t) = \mathbb{E}[X] + t\,\mathbb{E}[X^2] + \frac{t^2}{2!}\,\mathbb{E}[X^3] + \dots \quad \Longrightarrow M_X'(0) = \mathbb{E}[X]$$

$$M_X''(t) = \mathbb{E}[X^2] + t\,\mathbb{E}[X^3] + \frac{t^2}{2!}\,\mathbb{E}[X^4] + \dots \quad \Longrightarrow M_X''(0) = \mathbb{E}[X^2]$$

Theorem

If $M_X(t) < \infty$ in an interval around $t = 0$, then $M_X^{(k)}(0) = \mathbb{E}[X^k]$.

The MGF determines the distribution

Similarly to the PMF/PDF and CDF, the MGF also uniquely characterizes the distribution of a random variable.

Theorem

Let X and Y be random variables such that $M_X(t) = M_Y(t) < \infty$ in an interval around $t = 0$. Then they have the same distribution.

Implication: We may derive the distribution of a statistic by computing its MGF.

For sums of independent random variables $X_1, \ldots, X_n$, this is usually *easier* than computing the PDF or CDF, because

$$\begin{aligned} M_{X_1+\ldots+X_n}(t) &= \mathbb{E}[e^{t(X_1+\ldots+X_n)}] \\ &= \mathbb{E}[e^{tX_1}] \times \ldots \times \mathbb{E}[e^{tX_n}] = M_{X_1}(t) \times \ldots \times M_{X_n}(t) \end{aligned}$$

Sum of IID Poisson variables

Example: Let $X_1, \ldots, X_n \overset{IID}{\sim} \text{Poisson}(\lambda)$. What is the distribution of $X_1 + \ldots + X_n$?

$$M_{X_1+\ldots+X_n}(t) = \mathbb{E}\left[e^{t(X_1+\ldots+X_n)}\right]$$

$$= \mathbb{E}\left[e^{tX_1} \times \ldots \times e^{tX_n}\right]$$

$$= \underbrace{\mathbb{E}\left[e^{tX_1}\right]}_{= M_{X_1}(t) = e^{\lambda(e^t - 1)}} \times \ldots \times \mathbb{E}\left[e^{tX_n}\right] = \underbrace{e^{n\lambda(e^t-1)}}_{\text{MGF of Poisson}(n\lambda)}$$

So $X_1 + \ldots + X_n \sim \text{Poisson}(n\lambda)$

Mean of IID normal variables

Example: Let $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(0,1)$. What is the distribution of the sample mean $\bar{X} = \frac{X_1 + \ldots + X_n}{n}$?

$$M_{\bar{X}}(t) = \mathbb{E}\left[e^{t\bar{X}}\right]$$

$$= \mathbb{E}\left[e^{t\left(\frac{X_1 + \ldots + X_n}{n}\right)}\right]$$

$$= \underbrace{\mathbb{E}\left[e^{\frac{t}{n}X_1}\right]}_{M_{X_1}\left(\frac{t}{n}\right) = e^{\frac{t^2}{2n^2}}} \times \ldots \times \mathbb{E}\left[e^{\frac{t}{n}X_n}\right] = \underbrace{e^{\frac{t^2}{2n}}}_{\text{MGF of } \mathcal{N}(0, \frac{1}{n})}$$

$$\Rightarrow \bar{X} \sim \mathcal{N}\left(0, \frac{1}{n}\right)$$

Sum of independent normal variables

Example: Let $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, n$ be independent normal random variables. For any constants $a_1, \ldots, a_n \in \mathbb{R}$, what is the distribution of $a_1 X_1 + \ldots + a_n X_n$?

$$M_{a_1X_1+\dots+a_nX_n}(t) = \mathbb{E}\left[e^{t(a_1X_1+\dots+a_nX_n)}\right]$$

$$= \underbrace{\mathbb{E}\left[e^{ta_1X_1}\right]}_{M_{X_1}(ta_1) = e^{ta_1\mu_1 + \frac{t^2a_1^2\sigma_1^2}{2}}} \times \mathbb{E}\left[e^{ta_2X_2}\right] \times \dots \times \mathbb{E}\left[e^{ta_nX_n}\right]$$

$$= e^{t(a_1\mu_1+\dots+a_n\mu_n) + \frac{t^2}{2}(a_1^2\sigma_1^2+\dots+a_n^2\sigma_n^2)}$$

$$\Rightarrow a_1X_1+\dots+a_nX_n \sim \mathcal{N}(a_1\mu_1+\dots+a_n\mu_n,\ a_1^2\sigma_1^2+\dots+a_n^2\sigma_n^2)$$

Derivation of χ_n^2 distribution

Example: Let $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(0,1)$. Last lecture, we called the distribution of $X_1^2 + \ldots + X_n^2$ the **chi-squared distribution** χ_n^2, and claimed this is a special case of the Gamma distribution. Why?

Consider $n=1$: X_1^2

$$M_{X_1^2}(t) = \mathbb{E}[e^{tX_1^2}]$$

$$= \int_{-\infty}^{\infty} e^{tx^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$= \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi}} e^{(t-\frac{1}{2})x^2}}_{\frac{1}{\sqrt{1-2t}} \cdot \underbrace{\sqrt{\frac{1-2t}{2\pi}} e^{-\frac{1}{2}(1-2t)x^2}}_{\text{PDF of } \mathcal{N}(0, \frac{1}{1-2t})}} dx \qquad = \infty \text{ if } t \geq \frac{1}{2}$$

MGF of Gamma$(\frac{1}{2}, \frac{1}{2})$ $\rightsquigarrow$

$$= \begin{cases} \infty & \text{if } t \geq 1/2 \\ \frac{1}{(1-2t)^{1/2}} & \text{if } t < 1/2 \end{cases}$$

$\Rightarrow X_1^2 \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$

Derivation of χ_n^2 distribution

Consider $X_1^2 + \dots + X_n^2$:

$$M_{X_1^2+\dots+X_n^2}(t) = \mathbb{E}\left[e^{t(X_1^2+\dots+X_n^2)}\right]$$

$$= \underbrace{\mathbb{E}\left[e^{tX_1^2}\right]}_{= \begin{cases} \infty & \text{if } t \geq \frac{1}{2} \\ \frac{1}{(1-2t)^{1/2}} & \text{if } t < 1/2 \end{cases}} \times \dots \times \mathbb{E}\left[e^{tX_n^2}\right]$$

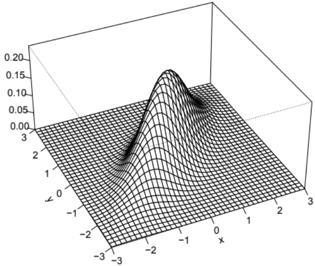
$$= \begin{cases} \infty & \text{if } t \geq 1/2 \\ \frac{1}{(1-2t)^{n/2}} & \text{if } t < 1/2 \end{cases} \leftarrow \text{MGF of Gamma}\left(\tfrac{n}{2}, \tfrac{1}{2}\right)$$

$$\Rightarrow X_1^2 + \dots + X_n^2 \sim \text{Gamma}\left(\tfrac{n}{2}, \tfrac{1}{2}\right)$$

The multivariate normal distribution

The multivariate normal distribution

The multivariate normal distribution in $\mathbb{R}^k$ is a joint distribution of k continuous random variables $(X_1, \ldots, X_k)$, which generalizes the normal distribution for a single variable $k = 1$.



It is parametrized by a **mean vector** $\boldsymbol{\mu} \in \mathbb{R}^k$ and a symmetric **covariance matrix** $\boldsymbol{\Sigma} \in \mathbb{R}^{k \times k}$, and we write the distribution as

$$(X_1, \ldots, X_k) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

The standard multivariate normal

An important example is the **standard multivariate normal** distribution in $\mathbb{R}^k$, which describes the joint distribution of

$$X_1, \ldots, X_k \overset{IID}{\sim} \mathcal{N}(0,1)$$

We denote the standard multivariate normal by

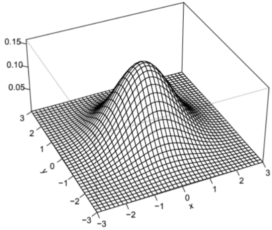
$$(X_1, \ldots, X_k) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

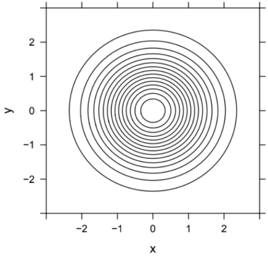
with mean vector $\mathbf{0} \in \mathbb{R}^k$ and identity covariance matrix $\mathbf{I} \in \mathbb{R}^{k \times k}$.

This distribution has joint PDF

$$f(x_1, \ldots, x_k) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2} = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}(x_1^2 + \ldots + x_k^2)}$$

Symmetry of the standard multivariate normal





Observe that this joint PDF

$$f(x_1, \ldots, x_k) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}(x_1^2 + \ldots + x_k^2)}$$

depends on $\mathbf{x} = (x_1, \ldots, x_k)$ only via its *length* $\sqrt{x_1^2 + \ldots + x_k^2}$.

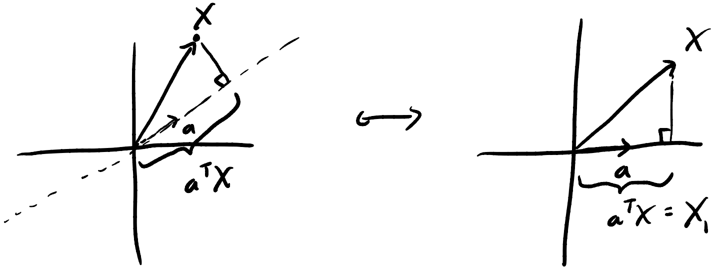
Thus the standard multivariate normal distribution is *symmetric* with respect to rotations/reflections about the origin!

Consequences of symmetry

Let $\mathbf{X} = (X_1, \ldots, X_k) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and let $\mathbf{a} \in \mathbb{R}^k$ be any vector having length 1. Then

$$\mathbf{a}^\top \mathbf{X} \sim \mathcal{N}(0, 1)$$

because this distribution is the same as for $(1, 0, \ldots, 0)^\top \mathbf{X} = X_1$.



Mean of IID normal variables

Example: Let $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(0,1)$. What is the distribution of the sample mean $\bar{X} = \frac{X_1 + \ldots + X_n}{n}$?

Let $a = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right) \in \mathbb{R}^n$

(Length of $a = \sqrt{\left(\frac{1}{\sqrt{n}}\right)^2 + \ldots + \left(\frac{1}{\sqrt{n}}\right)^2} = 1$)

So $a^T X \sim \mathcal{N}(0,1)$

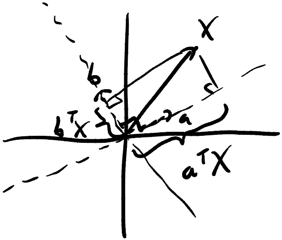
$\Rightarrow \bar{X} = \frac{1}{\sqrt{n}} \cdot a^T X \sim \mathcal{N}\left(0, \frac{1}{n}\right)$

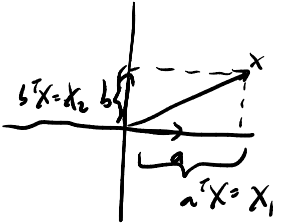
Consequences of symmetry

Let $\mathbf{X} = (X_1, \ldots, X_k) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$ be any two vectors having length 1 that are perpendicular to each other:

$$\mathbf{a}^\top \mathbf{b} = 0.$$

Then $\mathbf{a}^\top \mathbf{X}$ and $\mathbf{b}^\top \mathbf{X}$ are *independent* $\mathcal{N}(0, 1)$ random variables, because their joint distribution is the same as that of (X_1, X_2).





Independence of mean and residuals

Example: Let $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(0,1)$. What is the joint distribution of $\bar{X}$ and $X_1 - \bar{X}$?

Let $a = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$, length of $a = \frac{1}{\sqrt{n}}$

$b = \left(1 - \frac{1}{n}, -\frac{1}{n}, \ldots, -\frac{1}{n}\right)$, length of $b = \sqrt{1 - \frac{1}{n}}$

Importantly: $a^T b = 0$, so a and b are perpendicular.

So $\bar{X} = a^T X \sim \mathcal{N}(0, \frac{1}{n})$

$X_1 - \bar{X} = b^T X \sim \mathcal{N}(0, 1 - \frac{1}{n})$

} independent!