

Lecture 12

Recap on Gaussian distribution:

- We said if $X \sim N(\mu, \sigma^2)$ Then $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

- We also introduced $Z \sim N(0, 1)$ (standard normal distribution)

- We showed that if $X \sim N(\mu, \sigma^2)$, Then $\frac{X-\mu}{\sigma} \sim N(0, 1)$

- This comes from the fact that $aX+b \sim N(a\mu+b, a^2\sigma^2)$ for $a, b \in \mathbb{R}$
(In the above $a = \frac{1}{\sigma}$ $b = -\frac{\mu}{\sigma}$)

- This is useful because

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$\uparrow$ cdf of $N(0, 1)$

Move to bivariate/multivariate normal

Suppose Z_1 and Z_2 are independent standard normal r.v.s

Define the constants $\mu_1, \mu_2 \in \mathbb{R}$; $\sigma_1, \sigma_2 > 0$; and $\rho \in (-1, 1)$

Given this, we will define two new r.v.s:

$$X_1 = \sigma_1 Z_1 + \mu_1$$

$$X_2 = \sigma_2(\rho Z_1 + \sqrt{1-\rho^2} Z_2) + \mu_2$$

Claim: The joint pdf of (X_1, X_2) is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\} (*)$$

$$\exp(x) \equiv e^x$$

Proof: Exercise (use functions of bivariate r.v.s)

Note: (X_1, X_2) follow a bivariate normal distribution

As with univariate normal distributions, we use similar notation

$$(X_1, X_2) \sim N(\mu, \Sigma)$$

$\uparrow$ this is a 2×1 vector
 $\uparrow$ this is 2×2 matrix

In general Suppose $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$ we have $\sigma_{12} = \sigma_{21}$

and $X \equiv (X_1, X_2)$ with $X \sim N(\mu, \Sigma)$, Then

$$X_1 \sim N(\mu_1, \sigma_1^2)$$

$$X_2 \sim N(\mu_2, \sigma_2^2)$$

Notes on bivariate normal

- Based on $*$ we have $X_1 \sim N(\mu_1, \sigma_1^2)$ $E[X_1] = \mu_1$; $\text{var}(X_1) = \sigma_1^2$
 $X_2 \sim N(\mu_2, \sigma_2^2)$ $E[X_2] = \mu_2$; $\text{var}(X_2) = \sigma_2^2$
 $\text{Cor}(X_1, X_2) = \rho$

- X_1 and X_2 are independent if and only if $\rho = 0$

(Recall That $\rho = 0$ does not always imply independence except for this case)

✓ $a_1, a_2, b \in \mathbb{R}$

- If $Y = a_1 X_1 + a_2 X_2 + b$, Then $Y \sim N(a_1 \mu_1 + a_2 \mu_2 + b, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + 2a_1 a_2 \rho \sigma_1 \sigma_2)$

Exercise

- If $X_1 \perp\!\!\!\perp X_2$, Then $\rho = 0$ and this simplifies to $N(a_1 \mu_1 + a_2 \mu_2 + b, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)$

Exercise

- You can show using functions of bivariate r.v.s

Show This using m.g.f.s

If $X \sim N(\mu, \sigma^2)$ Then $M_X(t) = e^{\mu t + \frac{1}{2} \sigma^2 t^2}$

$$\begin{aligned} \underbrace{M_{a_1 X_1 + a_2 X_2 + b}}_Y(t) &= \mathbb{E}[\exp((a_1 X_1 + a_2 X_2 + b)t)] \\ &= \mathbb{E}[\exp(a_1 X_1 t + a_2 X_2 t + bt)] \\ &= \mathbb{E}[\exp(a_1 X_1 t) \exp(a_2 X_2 t) \exp(bt)] \\ &= \underbrace{\mathbb{E}[\exp(a_1 X_1 t)] \mathbb{E}[\exp(a_2 X_2 t)]}_{\text{(Independence)}} \exp(bt) \end{aligned}$$

(Independence)

$$\begin{aligned} &= M_{X_1}(a_1 t) M_{X_2}(a_2 t) \exp(bt) \\ &= \exp\left(\mu_1(a_1 t) + \frac{\sigma_1^2}{2}(a_1 t)^2\right) \exp\left(\mu_2(a_2 t) + \frac{\sigma_2^2}{2}(a_2 t)^2\right) \exp(bt) \\ &= \exp\left(\underbrace{a_1 \mu_1 + a_2 \mu_2 + b}_Y t + \frac{1}{2} \underbrace{(a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)}_Y\right) \\ &\text{This is the mgf of } Y \text{ where} \\ &Y \sim N(a_1 \mu_1 + a_2 \mu_2 + b, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2) \end{aligned}$$

- If (X_1, X_2) have the pdf in (*) Then

$$X_2 | X_1 = x_1 \sim N\left(\mu_2 + \rho \sigma_2 \left(\frac{x_1 - \mu_1}{\sigma_1}\right), (1 - \rho^2) \sigma_2^2\right)$$

Proof: exercise

Multivariate normal

DEF: If $X = (X_1, \dots, X_d)$ has a multivariate normal distribution with mean $\mu \in \mathbb{R}^d$ and covariance $\Sigma \in \mathbb{R}^{d \times d}$ Then we write $X \sim N(\mu, \Sigma)$ and X has joint pdf

$$f(x) = \frac{1}{(2\pi)^{d/2} \det(\Sigma)^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right] \quad x \in \mathbb{R}^d$$

If $d=2$, recover bivariate normal with mean $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \text{cov}(X_1, X_2) \\ \text{cov}(X_2, X_1) & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$\text{recall } \rho = \frac{\text{cov}(X_1, X_2)}{\sigma_1 \sigma_2}$$

$$\Rightarrow \text{cov}(X_1, X_2) = \rho \sigma_1 \sigma_2$$

Note: Σ is positive definite

$$\forall x \in \mathbb{R}, x^T \Sigma x > 0$$

by the spectral theorem $\Sigma = Q^T \Lambda Q$ and Σ is symmetric

Capital λ

orthonormal matrix

$$\text{diag}(\lambda_1, \dots, \lambda_d)$$

where eigenvalues of Σ (which are all positive)

We can also write $\Sigma = R R^T$ for some R and write $X = R Z + \mu$ where $Z \sim N(0, I)$

[univariate analog if $X \sim N(\mu, \sigma^2)$ then $\frac{X - \mu}{\sigma} \stackrel{d}{=} Z$
 $X \stackrel{d}{=} \sigma Z + \mu$]

$$R = \sqrt{\Sigma}$$

$$\sigma = \sqrt{\sigma^2}$$

Linear transformations of multivariate normals

If $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$ then $Y \equiv AX + b \sim N(A\mu + b, A\Sigma A^T)$ (~~AA~~)

[univariate analog if $X \sim N(\mu, \sigma^2)$ then $aX + b \sim N(a\mu + b, a^2 \sigma^2)$]
 $\uparrow$ m-variate normal
 $"A\Sigma A^T = A^2 \Sigma"$

In general, let $X = (X_1, \dots, X_d)$ be continuous with pdf f . Define $Y = (Y_1, \dots, Y_m)$ where $Y = AX$ for $A \in \mathbb{R}^{m \times d}$. Then the pdf of Y is given by

$$f_Y(y) = \frac{1}{\det(A)} f_X(A^{-1}y) \quad \text{for } y \in \mathbb{R}^m \quad Y = AX \Rightarrow X = A^{-1}Y$$

$\hookrightarrow$ We can use this to verify (~~AA~~) (or we can use mgfs)

The theorem for general transforms: Ch3 lecture notes

Stochastic convergence

Goal: we want to learn enough concepts/tools to prove:

- (LLN) ① (Weak) Law of Large Numbers "sample average $\approx$ true average in the long run"
- (CLT) ② Central Limit Theorem "some transform of sample average $\approx$ Normal for large samples"

To start, we'll define what " $\approx$ " mean. We'll do so through notions of stochastic convergence. We'll focus on two notions

- ① Convergence in probability (relevant for LLN)
- ② Convergence in distribution (relevant for CLT)

Convergence in probability

Consider a collection of r.v.s $\{X_n\}_{n=1}^{\infty}$. We say X_n converges in probability to a constant, $c \in \mathbb{R}$ if for all $t > 0$, we have that $\mathbb{P}(|X_n - c| > t) \rightarrow 0$ as $n \rightarrow \infty$.

• Shorthand for convergence in prob. $X_n \xrightarrow{P} c$

• Equivalently, $X_n \xrightarrow{P} c$ if $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - c| \leq t) = 1$ for all $t > 0$

Nonrandom case: $a_n \rightarrow a$ if $a-t < a_n < a+t$ for all $t > 0$ and all $n > N$ for some N
 Conv. in prob. case: $X_n \rightarrow c$ if $c-t < X_n < c+t$ for all $t > 0$ "with high probability"
 to see this, $\lim_{n \rightarrow \infty} P(|X_n - c| \leq t) = 1$ means that for arbitrary $\epsilon > 0$, there exists

$N \equiv N(\epsilon)$ such that for all $n > N$, we have that

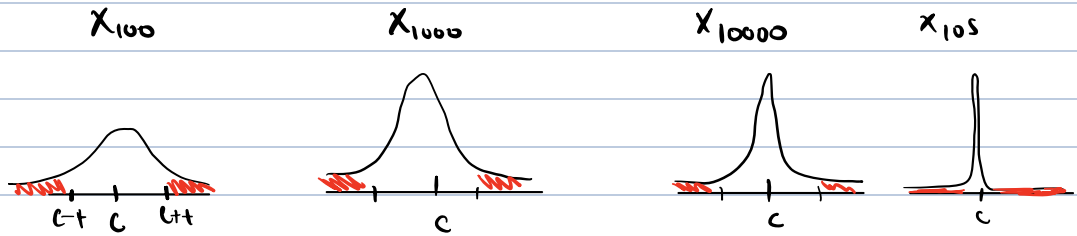
$$|P(|X_n - c| \leq t) - 1| < \epsilon$$

$$\Leftrightarrow 1 - \epsilon < P(|X_n - c| \leq t) < 1 + \epsilon$$

$$\Leftrightarrow 1 - \epsilon \leq P(c-t \leq X_n \leq c+t) < 1 + \epsilon$$

$$\Rightarrow P(c-t \leq X_n \leq c+t) \geq 1 - \epsilon$$

$\xrightarrow{P}$ looks like



THM: (Continuous mapping theorem). If $X_n \xrightarrow{P} c$ and g is continuous at c
 Then $g(X_n) \xrightarrow{P} g(c)$

Weak Law of Large Numbers

THM: If $X_1, \dots, X_n$ are iid with finite mean μ (and finite variance σ^2) Then we have that $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu$

$$\underbrace{\frac{1}{n} \sum_{i=1}^n X_i}_{\bar{X}} \rightarrow \mu \text{ mean } \forall \epsilon > 0, \exists N(\epsilon) \text{ such that } |\bar{X} - \mu| < \epsilon \quad \forall n > N$$

$$P(\lim_{n \rightarrow \infty} \bar{X} = \mu) = 1 \text{ (almost sure)}$$

Proof: Idea: we want to show that

$$P(|\bar{X} - \mu| > t) \rightarrow 0 \text{ for arbitrary } t > 0$$

$$\text{to do this, we often } P(|\bar{X} - \mu| > t) \leq \text{something} \rightarrow 0 \text{ (as } n \rightarrow \infty)$$

How do we find something? There are many inequalities of this form

- Markov's inequality
- Chebyshev's
- Chernoff's

We will use Chebyshev's inequality: In general, if X is a r.v. with $\text{var}(X) < \infty$ Then for any $t > 0$, we have that

$$P(|X - E[X]| > t) \leq \frac{\text{var}(X)}{t^2}$$

We want to apply Chebyshev to $\bar{X}$

(this is why we need $\sigma^2 < \infty$)

$$P(|\bar{X} - E[\bar{X}]| > t) \leq \frac{\text{var}(\bar{X})}{t^2}$$

$$E[\bar{X}] = \mu$$

$$\text{var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$\text{Then } P(|\bar{X} - \mu| > t) \leq \frac{\sigma^2}{nt^2} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Consequence of Chebyshev: In general, if you want to show that $Y_n \xrightarrow{P} y$, Then it is sufficient to show that

- ① $E[Y_n] \rightarrow y$
- ② $\text{var}(Y_n) \rightarrow 0$