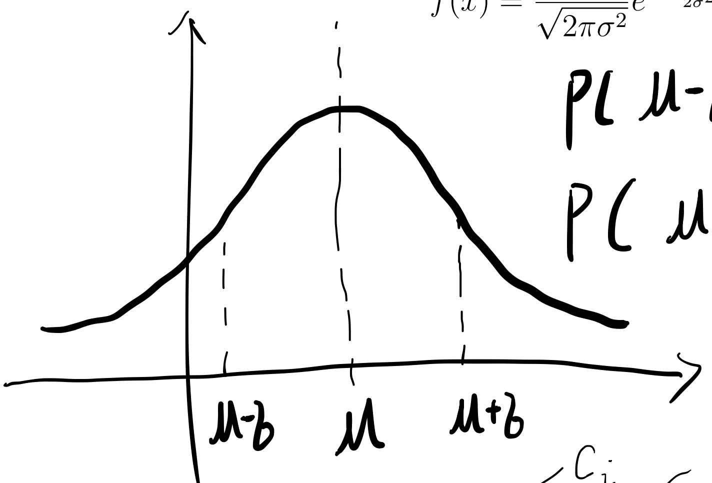
Packet 2: Functions of Random Variables

Chap 5.5 Random Variables related with Normal distributions

Normal distribution (Gaussian distribution) is originally found by observing that mean of sample offen follows a special bell shaped distribution.

$X \sim N(\mu, \sigma^2)$, $E(X) = \mu$, $Var(X) = \sigma^2$, has p.d.f. and m.g.f.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$



$P(\mu-\sigma < X < \mu+\sigma) = 0.68$

$P(\mu-2\sigma < X < \mu+2\sigma) = 0.95$

c_i $\quad$ c_i^2

Theorem 5.5-1: If X_1, X_2, ... X_n are independent random variables with $X_i \sim N(\mu_i, \sigma_i^2)$, then $Y = \sum_{i=1}^n c_i X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.

proof. $$M_Y(t) = \prod_{i=1}^n M_{X_i}(c_i t) = \prod_{i=1}^n e^{\mu_i c_i t + \frac{\sigma_i^2 c_i^2 t^2}{2}}$$

$$= e^{\underbrace{\sum_{i=1}^n c_i \mu_i}_{\text{mean}} \times t + \underbrace{\sum_{i=1}^n c_i^2 \sigma_i^2}_{\text{variance}} \times \frac{t^2}{2}}$$

is the m.g.f. for a normal r.v.

So $Y \sim N\left(\sum_{i=1}^n c_i \mu_i, \sum_{i=1}^n c_i^2 \sigma_i^2\right)$

If $c_i = \frac{1}{n}$ $\quad \mu_i = \mu \quad \sigma_i^2 = \sigma^2$

Corollary 5.5-1: If X_1, X_2, ... X_n are independent random variables with $X_i \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \sigma^2/n)$.

sample mean $\bar{X} \sim N\left(\sum_{i=1}^{n} \frac{1}{n}\mu, \ \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \sigma^2\right)$

$\sim N\left(\mu, \frac{1}{n}\sigma^2\right)$ $\quad$ as $n \to \infty$ $\quad Var(\bar{x}) \to 0$

Theorem 5.5-2: If X_1, X_2, ... X_n are independent random variables with $X_i \sim N(\mu, \sigma^2)$, then the sample mean $\bar{X}$ and the sample variance $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ are independent,

$$\frac{S^2(n-1)}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

$E\left(\frac{S^2(n-1)}{\sigma^2}\right) = n-1 \qquad \frac{(n-1)}{\sigma^2} E(S^2) = n-1 \qquad E(S^2) = \sigma^2$

sample variance S^2 is an unbiased estimator for true variance σ^2

Theorem 5.5-3: Student's t distribution $T = \frac{Z}{\sqrt{U/r}} \sim t(r)$, where $Z \sim N(0,1)$ and $U \sim \chi^2(r)$.
If X_1, X_2, ... X_n are independent random variables with $X_i \sim N(\mu, \sigma^2)$, then

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

$\bar{X} \sim N(\mu, \sigma^2/n) \qquad Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$

$U = \frac{S^2(n-1)}{\sigma^2} \sim \chi^2(n-1)$

$$T = \frac{Z}{\sqrt{U/(n-1)}} \sim t(n-1) \qquad \frac{\frac{\bar{X} - \mu}{\sqrt{\sigma^2}/\sqrt{n}}}{\sqrt{\frac{S^2(n-1)}{\sigma^2(n-1)}}} = \frac{\frac{\bar{X} - \mu}{\cancel{\sqrt{\sigma^2}}/\sqrt{n}}}{\frac{S}{\cancel{\sqrt{\sigma^2}}}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

T statistic

Chap 5.6 The Central Limit Theorem (CLT)

i.i.d. independently and identically distributed

CLT tells us that, with sufficiently many i.i.d. samples collected, the sample mean $\bar{X}$ follows $N(\mu, \sigma^2/n)$ approximately, regardless the true distribution of X_i.

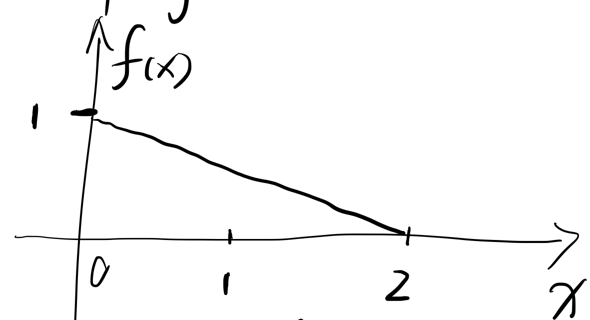
The distribution of X_i's is complicated or unknown.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \;\dot{\sim}\; N(\mu, \sigma^2/n) \quad \text{when } n \text{ is large}$$

where $E(X_i) = \mu \quad Var(X_i) = \sigma^2$

Example: $X_1 \cdots X_{18}$ i.i.d. with p.d.f.

$$f(x) = 1 - \frac{1}{2}x \qquad 0 < x < 2$$



Find the approximate distn of $\bar{X}$ by CLT

Sol: $\mu = E(X) = \int_0^2 x f(x)\, dx = 2/3$

$$E(X^2) = \int_0^2 x^2 f(x)\, dx = 2/3$$

$$\sigma^2 = Var(X) = E(X^2) - (E(X))^2 = 2/3 - (2/3)^2 = 2/9$$

by CLT $\quad \bar{X} \,\dot{\sim}\, N(\mu, \frac{\sigma^2}{n}) \quad N(\frac{2}{3}, \frac{2/9}{18}) \quad N(\frac{2}{3}, (\frac{1}{9})^2)$

* CLT is much easier than finding the exact distn of $\bar{X}$