Let $X_1, \ldots, X_n \sim$ iid X with $E[X] = 0$, $Var(X) = \sigma^2$. Let $Y = (X_1 + X_2 \cdots X_n)/n$.

Then, $M_Y(\lambda) \to e^{\frac{\lambda^2 \sigma^2}{2}}$

- MGF of Normal $(0, \sigma^2)$: $e^{\frac{\lambda^2 \sigma^2}{2}}$

- Y is said to converge in distribution to Normal $(0, \sigma^2)$

Using CLT to approximate probability

$$X_1, \ldots\ldots X_n \sim \text{iid } X$$

Let $\mu = E[X]$, $\sigma^2 = Var(X)$

$Y = X_1 + \ldots + X_n$

What is $P(Y - n\mu > \delta n\mu)$?

$\frac{Y - n\mu}{\sqrt{n}}$: approximately Normal$(0, \sigma^2)$

$$\frac{Y - n\mu}{\sqrt{n}\,\sigma} \approx \text{Normal}(0,1)$$

Divide by $\sqrt{n}\,\sigma$

Divide by $\sqrt{n}\sigma$

Date: __/__/__

- $F(z)$: CDF of Normal$(0,1)$
- $P(Y - n\mu > \delta n\mu) = P\left(\frac{Y-n\mu}{\sqrt{n}\sigma} > \frac{\delta\sqrt{n}\mu}{\sigma}\right)$

$$\approx 1 - F\left(\frac{\delta\sqrt{n}\mu}{\sigma}\right)$$

Linear Combination of Independent Normals

- $X_1, \ldots\ldots, X_n \sim$ independent normal
- Let $X_i \sim$ Normal(μ_i, σ_i^2)
- Suppose $Y = a_1X_1 + \ldots + a_nX_n$

Then, $Y \sim$ Normal(μ, σ^2)
where $\mu = E[Y] = a_1\mu_1 + \ldots\ a_n\mu_n$ and
$\sigma^2 = a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2$

Gamma Distribution

$$X \sim \text{Gamma}(\alpha, \beta)$$

if PDF $f_X(x) \propto x^{\alpha-1}e^{-\beta x}$, $x > 0$

- $\alpha > 0$: shape parameter, $\beta > 0$; rate parameter
  $\theta = 1/\beta$: scale parameter
- Mean $\Rightarrow \frac{\alpha}{\beta}$

Date: __/__/____

- Variance : $\frac{\alpha}{\beta^2}$
- MGF : $\left(1-\frac{\lambda}{\beta}\right)^{-\alpha}$, $\lambda < \beta$
- Sum of n iid $Exp(\beta)$ is $Gamma(n, \beta)$
- square of $Normal(0, \sigma^2)$ is Gamma $\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right)$

Cauchy distribution

$$X \sim Cauchy(\theta, \alpha^2)$$

if PDF $f_X(x) = \frac{1}{\pi} \frac{\alpha}{\alpha^2 + (x-\theta)^2}$

- θ : location parameter, $\alpha > 0$: scale parameter
- Mean, variance & MGF undefined
- Suppose $X, Y \sim$ iid $Normal(0, \sigma^2)$. Then, $\frac{X}{Y} \sim Cauchy(0, 1)$

$X \sim \text{Beta}(\alpha, \beta)$

if PDF $f_X(x) \propto x^{\alpha-1}(1-x)^{\beta-1}$, $0<x<1$

- $\alpha > 0$, $\beta > 0$: shape parameters
- ~~Mean~~ Mean $= \dfrac{\alpha}{\alpha+\beta}$, Variance $= \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Beta$(\alpha, 1)$ has PDF $\propto x^{\alpha-1}$: power function distribution

- Suppose $X \sim \text{Gamma}(\alpha, 1/\theta)$, $Y \sim \text{Gamma}(\beta, 1/\theta)$, then

$$\frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta)$$

Date: __/__/____

Distribution of Sample Mean

$X_1, \ldots\ldots, X_n \sim$ iid Normal (μ, σ^2)

- $\bar{X} = \frac{1}{n} X_1 + \ldots\ldots + \frac{1}{n} X_n$
- Sample mean is a linear combination of iid normal random variables.

$$\bar{X} \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$$

- $E[\bar{X}] = \mu$
- $Var(\bar{X}) = \frac{\sigma^2}{n}$

Sum of squares of normal squares: Chi-square

$X_1, \ldots\ldots X_n \sim$ iid Normal $(0, \sigma^2)$

- X_i^2 : Gamma $\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right)$, independent

Sum of n independent Gamma (α, β) is Gamma $(n\alpha, \beta)$

Date: __/__/____

- $\text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$: called Chi-square distribution with n degrees of freedom, denoted χ_n^2

Sample mean and variance of normal samples

Suppose $X_1, \dots, X_n \sim \text{Normal}(\mu, \sigma^2)$. Then,

- $\bar{X} \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$, Chi-square with $n-1$ degrees of freedom
- $\bar{X}$ and S^2 are independent.

Date: __/__/____

Parameter Estimation

- iid samples

  $X_1, X_2 \ldots X_n \sim \text{iid } X$

- X has a distribution described by some parameters $\theta_1, \theta_2, \ldots$
  - Parameters take real values, $\theta_i \in \mathbb{R}$

- Parameter estimation: What is θ_1? What is θ_2?

- Estimator for a parameter θ.
  - function of the samples: $\hat{\theta}(X_1, \ldots X_n)$
  - Notation: $\hat{\theta}$ is an estimator for parameter θ

- Parameter vs estimator
  - θ : constant parameter, not random variable
  - $\hat{\theta}$: function of n random variables, $\therefore$ it is a random variable

  $\hat{\theta}$ will have distribution, PMF or PDF

Date: __/__/____

Estimation Error

- θ : parameter, $\hat{\theta}(X_1, X_2 \dots X_n)$: estimator
  - Error : $\hat{\theta}(X_1, X_2 \dots X_n) - \theta$ is a random variable.

- We expect the estimator random variable $\hat{\theta}(X_1, X_2, \dots X_n)$ to take values around the actual parameter θ. So, the random variable 'Error' should take values close to 0.
  - How to express this mathematically?
    $P(|\text{Error}| > \delta)$ should be small.

- Chebyshev bound :

$$P\left(\left|\text{Error} - E[\text{Error}]\right| > \delta\right) \leq \frac{\text{Var}(\text{Error})}{\delta^2}$$

- Good design principles
  - $E[\text{Error}]$ should be close to or equal to 0.
  - $\text{Var}(\text{Error}) \to 0$ with n

Date: __/__/____

Bias

The bias of the estimator $\hat{\theta}$ for a parameter θ, denoted by $Bias(\hat{\theta}, \theta)$ is defined as

$$Bias(\hat{\theta}, \theta) = E[\hat{\theta}] - \theta$$

- Since Error $= \hat{\theta} - \theta$, bias is the expected value of Error.

- An estimator with bias equal to 0 is said to be an unbiased estimator.

Risk

The (squared-error) risk of the estimator $\hat{\theta}$ for a parameter θ, denoted $Risk(\hat{\theta}, \theta)$ is defined as

$$Risk(\hat{\theta}, \theta) = E[(\hat{\theta} - \theta)^2]$$

- Also known as MSE : mean squared error

- Squared-error risk is the second moment of error

Date: __/__/____

Variance

$$Var(\hat{\theta}) = E\left[(\hat{\theta} - E[\hat{\theta}])^2\right]$$

- $Var(\text{Error}) = Var(\hat{\theta})$

Bias- Variance trade-off

$$Risk(\hat{\theta}, \theta) = Bias(\hat{\theta}, \theta)^2 + Var(\theta)$$

i.e.

$$E\left[(\hat{\theta} - \theta)^2\right] = E[\hat{\theta} - \theta]^2 + E\left[(\hat{\theta} - E[\hat{\theta}])^2\right]$$

$X_1, \ldots\ldots, X_n \sim \text{iid } X$

- Sample moments
$$M_k(X_1, \ldots\ldots X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

- One sampling instance: $x_1, \ldots\ldots, x_n$
  - 1st sample moment : $m_1 = \frac{1}{n}(x_1 + \ldots + x_n)$
  - k-th sample moment : $m_k = \frac{1}{n}(x_1^k + \ldots + x_n^k)$

- M_k is a random variable, and m_k is the value taken by it one sampling instance.
  - If sampling is repeated, the random variable M_k will take different values
  - We expect that M_k will take values around $E[X^k]$

Date: __/__/___

Method of moments

- Procedure
  - Equate sample moments to expression for moments in terms of unknown parameters
  - Solve for the unknown parameters

- One parameter θ usually needs one moment:
  - Sample moment: m_1
  - Distribution moment: $E[X] = f(\theta)$
  - Solve for θ from $f(\theta) = m_1$ in terms of m_1
  - $\hat{\theta}$: replace m_1 by M_1 in above ~~conditi~~ solution

- Two parameters θ_1, θ_2 usually needs two moments
  - Sample moments: m_1, m_2
  - Distribution moment: $E[X] = f(\theta_1, \theta_2)$, $E[X^2] = g(\theta_1, \theta_2)$
  - Solve for θ_1, θ_2 from $f(\theta_1, \theta_2) = m_1$ & $g(\theta_1, \theta_2) = m_2$ in terms of m_1, m_2
  - $\hat{\theta}_1, \hat{\theta}_2$: replace m_1 by M_1 and m_2 by M_2 in above solution

$$L(x_1, \ldots x_n) = \prod_{i=1} f_X(x_i; \theta_1, \theta_2 \ldots)$$

Maximum Likelihood (ML) estimator

$$\theta_1^*, \theta_2^*, \ldots = \arg \max_{\theta_1, \theta_2} \prod_{i=1}^{n} f_X(x_i; \theta_1, \theta_2 \ldots)$$

- Find parameters that maximize likelihood for a given set of samples

Date: __/__/____

Consistency of estimators

$$X_1, \ldots\ldots X_n \sim \text{iid } f_X(x;\theta)$$

- Estimator : $\hat{\theta}$, Error $= \hat{\theta} - \theta$
- 'error' is a random variable. As n increases, we expect 'error' to take values that are close to zero.

$$P(|\text{error}| > \delta) \to 0 \quad \text{as } n \to \infty \text{ for any } \delta > 0$$

- If an estimator satisfies the above requirement, it is said to be consistent.
- Technically, the above requirement is called convergence in probability.

- Method of moments estimators
  - If parameter is mean or variance, they will be unbiased.
  - For most other cases, they may be biased.
  - Usually, consistent
    - sample moments converge to distribution moments
    - If MME estimate is a continuous function of moments, then the estimate converges too

- Maximum likelihood estimators
  - Consistent
  - Bias vanishes in a limiting sense with growing n

Date: __/__/____

Estimation of sample mean and confidence interval

- Estimator : $\hat{\mu} = \frac{X_1 + X_2 + \dots + X_n}{n}$

- Confidence interval
  - Suppose $P(|\hat{\mu} - \mu| < \alpha) = \beta$, where α is a small fraction and β is a large fraction
  - $\hat{\mu}$ is one sampling instance: estimate with margin of error $(100\alpha)\%$ at confidence level $(100\beta)\%$.

- How to find α, β for which $P(|X - \mu| < \alpha) = \beta$?
  - Suppose X is continuous and has CDF F_X
  - ~~PDF~~ $P(X \leq x) = F_X(x)$

$$P(|X - \mu| < \alpha) = P(\mu - \alpha < X < \mu + \alpha)$$
$$= F_X(\mu + \alpha) - F_X(\mu - \alpha)$$

  - Given β, find α such that $F_X(\mu + \alpha) - F_X(\mu - \alpha) = \beta$

Date: __/__/____

How to find confidence interval when standard deviation is known?

Step 1: List n, $\bar{x}$, σ

Step 2: Get z-score for the given confidence %.

Step 3 $CI \rightarrow \bar{X} \pm \frac{Z\sigma}{\sqrt{n}}$

How to find confidence interval when standard deviation is unknown?

Step 1: Calculate $\alpha = \frac{1 - \text{confidence}}{2}$

Step 2: Get t-score for $(n-1)$ degrees of freedom and α.

Step 3: $CI \rightarrow \bar{X} \pm T_{score} \frac{\sigma}{\sqrt{n}}$

Confidence Interval for Binomial's 'p'

$$CI \Rightarrow \hat{p} \pm Z\sqrt{\frac{\hat{p}\hat{q}}{n}}$$