What to study today (Oct 5, 2020)?

2. Probability and Distribution (Chp 1-3)

2.1 Probability (Chp1.1-4)

2.2 Random Variable and Distribution (Chp1.5-10)

2.3 Multivariate Distribution (Chp2)

- *2.3.1 Basic Concepts with Two Random Variables*
- *2.3.2 Conditional Distribution and Expectation*
- **2.3.3 Extension to Several Random Variables**

2.4 Some Important Distributions (Chp3)

- **2.4.1 Discrete Distributions**
- **2.4.2 Continuous Distributions**
- *2.4.3 Multivariate Distributions*
- *2.4.4 Distributions Induced from Others*

2.3.3 Extension to Several Random Variables: General Issues

Consider K (> 2) rvs $X_1, X_2, \ldots, X_K$: (*We have to miss a lot if studing them one at a time or two at a time.*)

Definition. The **joint cdf** of the random vector $(X_1, X_2, \ldots, X_K)$ is $F(x_1, x_2, \ldots, x_k) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_K \leq x_k)$ for $-\infty < x_1, x_2, \ldots, x_K < \infty$.

- In general $X_1 \sim F_{X_1}(x_1) = F(x_1, \infty, \ldots, \infty)$, $X_1, X_2 \sim F_{X_1,X_2}(x_1, x_2) = F(x_1, x_2, \ldots, \infty)$, etc.
- When $X_1, \ldots, X_K$ are discrete, the **joint pmf** of $(X_1, \ldots, X_K)$ is $p(x_1, \ldots, x_K) = P(X_1 = x_1, \ldots, X_K = x_K)$; when $X_1, \ldots, X_K$ are continuous, the **joint pdf** of $(X_1, \ldots, X_K)$ is $f(x_1, \ldots, x_K)$ such that $P\big((X_1, \ldots, X_K) \in A\big) = \int \ldots \int_A f(x_1, \ldots, x_K) dx_1 \ldots dx_K$ for $A \in \mathcal{R}^K$.
- K rvs $X_1, \ldots, X_K$ are **independent** iff $F(x_1, \ldots, x_K) = F_{X_1}(x_1) \ldots F_{X_K}(x_K)$ for $-\infty < x_1, \ldots, x_K < \infty$.
- If $Y = g(X_1, \ldots, X_K)$, $E(Y) = \int \ldots \int_{\mathcal{R}^K} g(x_1, \ldots, x_K) dF(x_1, \ldots, x_K)$.

2.3.3 Extension to Several Random Variables: Linear Combination

Consider linear combinations of rvs $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$: $T = \sum_{i=1}^n a_i X_i$ and $W = \sum_{j=1}^m b_j Y_j$.

- $E(T) = \sum_{i=1}^n a_i E(X_i)$; $E(W) = \sum_{j=1}^m b_j E(Y_j)$
- $V(T) = \sum_{i=1}^n a_i^2 V(X_i) + 2\sum_{i<j} a_i a_j Cov(X_i, X_j)$
- $Cov(T, W) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(X_i, Y_j)$.

In the special case with rvs $X_1, \ldots, X_n$ indept and identically distributed (iid) and $E(X_i) = \mu$, $Var(X_i) = \sigma^2$,

- $E(T) = \left[\sum_{i=1}^n a_i\right]\mu$,
- $V(T) = \left[\sum_{i=1}^n a_i^2\right]\sigma^2$,
- Moreover, the mgf of T is $M(u) = \prod_{i=1}^n M_X(a_i u)$.
- If $X_1, \ldots, X_n$ are indpt of $Y_1, \ldots, Y_m$, $Cov(T, W) = 0$.

2.4 Some Important Distributions (Chp3)

2.4.1 Discrete Distributions: Discrete Uniform Distribution

Definition. r.v. X has a **discrete uniform** distribution on $a_1, \ldots, a_m$, if

$$p(x) = 1/m, \quad x = a_1, \ldots, a_m.$$

Physical Setting: X takes each of its possible values equally likely.

- $E(X) = (a_1 + \ldots + a_m)/m = \sum_{i=1}^{m} a_i/m$; $Var(X) = \sum_{i=1}^{m}(a_i - E(X))^2/m$.
- Example. X is the outcome attained by rolling a fair six-sided die: $p(x) = 1/6, \quad x = 1, \ldots, 6$.

2.4.1 Discrete Distributions: Binomial Distributions and Related

Definition. A random experiment is called a **Bernoulli experiment** if it has two possible outcomes, say, success vs failure.

- Define rv as $X(success) = 1$ and $X(failure) = 0$. Then X is a **Bernoulli random variable**.
- The distribution of X is called the **Bernoulli distribution**. Its pmf is $p(x) = \theta^x(1-\theta)^{1-x},\ for\ x = 0, 1$ if $P(X = 1) = \theta$.
  - $E(X) = \theta$; $Var(X) = \theta(1-\theta)$.
  - The mgf of X is $M(t) = 1 - \theta + e^t\theta$.

Definition. When a Bernoulli experiment is repeated n times *independently*, a sequence of n **Bernoulli trials** occurs.

2.4.1 Discrete Distributions: Binomial Distributions and Related

Definition. The distribution of a rv X is called a **binomial** distribution if its pmf is $p(x) = P(X = x) = \left(\begin{array}{c} n \\ x \end{array} \right) \theta^x (1-\theta)^{n-x}$ for $x = 0, 1, \ldots, n$, denoted by $X \sim B(n, \theta)$.

Physical Setting: Consider n Bernoulli trials, where the probability of success in every trial is θ. The the distribution of rv $X=$ the number of successes is $B(n, \theta)$.

- $E(X) = n\theta$; $Var(X) = n\theta(1-\theta)$
- Example. Flipping an even coin three times independently. $X=$number of heads. $X \sim B(3, 1/2)$.
- The Bernoulli distribution is $B(1, \theta)$.
- If $X \sim B(n, \theta)$, then $X = Y_1 + Y_2 + \cdots + Y_n$ with $Y_1, \ldots, Y_n$ indpt and $P(Y_1 = 1) = \cdots = P(Y_n = 1) = \theta$.

2.4.1 Discrete Distributions: Binomial Distributions and Related

Physical Setting: Consider a sequence of Bernoulli trials with probability θ of success. Let X denote the trial number on which the first success occurs.

Definition. The distribution of rv X is a **geometric** distribution with pmf

$$p(x) = P(X = x) = \theta(1-\theta)^{x-1}, \quad x = 1, 2, \ldots$$

with $0 \leq \theta \leq 1$.

- $E(X) = \frac{1}{\theta}$; $Var(X) = \frac{1-\theta}{\theta^2}$.
- Example. Toss an even coin until a head. The number of attempts follows the geometric distribution with $\theta = 1/2$.

2.4.1 Discrete Distributions: Binomial Distributions and Related

Physical Setting: Consider a sequence of Bernoulli trials with probability of success θ. Let X denote the trial number on which the rth success occurs.

Definition. The distribution of X is a **negative binomial** distribution with its pmf

$$p(x) = P(X = x) = \begin{pmatrix} x-1 \\ r-1 \end{pmatrix} \theta^r (1-\theta)^{x-r}, \quad x = r, r+1, \ldots$$

with $r \geq 0$ and $0 \leq \theta \leq 1$. Denote $X \sim NB(r, \theta)$.

- $E(X) = \frac{r}{\theta}$; $Var(X) = \frac{r(1-\theta)}{\theta^2}$.
- Example. Toss an even coin until the 3rd head. The number of attempts follows $NB(r, \theta)$ with $\theta = 1/2$ and $r = 3$.

2.4.1 Discrete Distributions: Hypergeometric Distribution.

Definition. r.v. X has a **hypergeometric** distn if

$$p(x) = P(X = x) = \frac{\begin{pmatrix} N_1 \\ x \end{pmatrix} \begin{pmatrix} N_2 \\ n - x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

for $\max(0, n - N_2) \leq x \leq \min(n, N_1)$ with $N = N_1 + N_2$.

Physical Setting: Randomly select n items without replacement from a group of $N = N_1 + N_2$ items, where N_1 items are in Category 1 and N_2 in Category 2. Let X be the number of selected items in Category 1.

2.4.1 Discrete Distributions: Poisson Distribution.

Definition. A r.v. X has a **Poisson** distribution, denoted by $X \sim Poisson(\lambda)$, if its pmf is

$$P(X = x) = p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$$

The distn is named after S.D. Poisson (1781-1840).

Comments.

- The Poisson distn is especially good at modelling rare events.
- $P(X = 0) = e^{-\lambda}$; $E(X) = Var(X) = \lambda$.
- $X \sim Poisson(\lambda)$ vs $X \sim Bin(n, \theta)$: difference and connection?
- Consider $X_1 \sim Poisson(\lambda_1)$ and $X_2 \sim Poisson(\lambda_2)$. If $X_1 \perp\!\!\!\perp X_2$, $Y = X_1 + X_2 \sim Poisson(\lambda)$ with $\lambda = \lambda_1 + \lambda_2$.

2.4.2 Continuous Distributions: Uniform Distribution.

Definition. A rv X has a Uniform(a,b) distribution if its pdf is

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & otherwise \end{cases}$$

- $E(X) = (a+b)/2$; $Var(X) = (b-a)^2/12$.
- Special case: $X \sim U(0,1)$.

2.4.2 Continuous Distributions: Normal Distribution

The most important distribution in all of Statistics is the normal (Gaussian) distribution.

Definition. A r.v. X has a *normal* distribution if its pdf

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right\}, \quad -\infty < x < \infty,$$

where $\sigma > 0$. Denote it by $X \sim N(\mu, \sigma^2)$.

- If $X \sim N(\mu, \sigma^2)$, $E(X) = \mu$ and $V(X) = \sigma^2$.
- If $X \sim N(\mu, \sigma^2)$, $f(x) > 0$ for all x and the cdf $F(x) = \int_{-\infty}^{x} f(u)du$ has no closed form.
- $N(\mu, \sigma^2)$: a family of distributions.
  - e.g. $N(0,1)$, *the standard normal distribution*. $F(x)$ of $N(0,1)$ is often denoted by $\Phi(x)$ and the rv by Z.

Proposition. If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

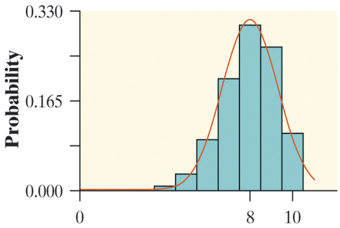
Very Useful!

Example: The number of hours that people watch TV is normally distributed with mean 6.0 hours and standard deviation 2.5 hours. (Is this reasonable?) What is the probability that a randomly selected person watches more than 8 hours of TV per day? [.2119]

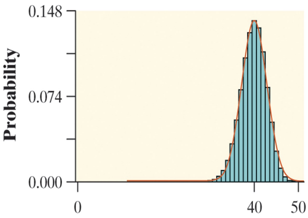
Recall that, if $X \sim B(n, \theta)$, its pmf is

$$P(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, \ldots, n.$$ When $n >> 1$, it is hard to calculate associated quantities in general.

As n gets larger, something interesting happens to the shape of a binomial distribution $B(n, \theta)$:



$n = 10, p = 0.8$



$n = 50, p = 0.8$

Proposition. Consider r.v. $X \sim B(n, p)$ where $np \geq 5$ and $n(1-p) \geq 5$. Then $X \sim N(\mu, \sigma^2)$ with $\mu = np, \sigma^2 = np(1-p)$.

2.4.2 Continuous Distributions: Exponential Distribution

Definition. A r.v. X has an **exponential** distribution with $\lambda > 0$, denoted by $X \sim \mathit{Exponential}(\lambda)$ or $\mathit{NE}(\lambda)$ if it has pdf

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

- The pdf is decreasing for $x > 0$, and asymmetric.
- The cdf is $F(x) = 1 - e^{-\lambda x}$ for $x > 0$.
- $E(X) = 1/\lambda$ and $V(X) = 1/\lambda^2$.
- $\mathit{NE}(\lambda)$ is a special case, when $\alpha = 1, \beta = 1/\lambda$, of the Gamma distribution $\mathit{Gamma}(\alpha, \beta)$:

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha} \Gamma(\alpha)}, \quad x > 0,$$

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ and $\alpha > 0, \beta > 0$.

More on the exponential distn ... ...

- The exponential distribution has the *memoryless* property:

$$P(X > a + b \big| X > a) = P(X > b), \quad a > 0, \; b > 0.$$

e.g. Suppose that the lifespan of a lightbulb in hours $X \sim NE(\lambda)$. The prob of a used lightbulb (that has already lasted a hours) lasts an additional b hours or more is the same as a new lightbulb does.

- Recall $\{N(t), t > 0\}$ is a Poisson process with the rate of λ, (the number of events over $[0, 1]$) $X = N(1) \sim Poisson(\lambda)$. In fact, $N(t) \sim Poisson(\lambda t)$.
  Y=the waiting time until the first event follows $NE(\lambda)$:

$$P(Y \leq y) = 1 - P\big(N(y) = 0\big) = 1 - e^{-\lambda y}.$$

What will we do in the next class?

1. *Introduction*

2. *Probability and Distribution (Chp 1-3)*

2.4 Some Important Distributions (Chp3)

- *2.4.1 Discrete Distributions*
- *2.4.2 Continuous Distributions*
- **2.4.3 Multivariate Distributions**
- **2.4.4 Distributions Induced from Others**

3. *Essential Topics in Mathematical Statistics (Chp 4-6)*

4. *Further Topics, Selected from Chp 7-11*