EE2012/ST2334 Discussion Points

Liu Zichen

May 1, 2019

1. **[sample space]** The color of a single pixel on the screen can usually be controlled by R, G, B components. Suppose the intensity value of R, G, B can only be taken from $\{0, 1, 2\}$.
   (a) What's the sample space of the color of a single pixel?
   (b) If a low-resolution image consists of 28×28 pixels, what's the size of sample space for such an image?
2. **[event probability]** Think of an event that has probability $p(x) = \frac{\pi}{4}$.
3. **[Law]** Write down the De Morgans Law for 3 event case.
4. **[counting]** How the multiplication principle and addition principle can be represented using a tree diagram? Come up with an example to show a $(2 \times 6 \times 2)$ and a $(2 + 3)$ case.
5. **[permutation]** $P(n, k) = \frac{n!}{(n-k)!}$. Explain the equation on the left. Does the order matter in permutation?
   When do we use $(n-1)!$ to calculate permutation?
   When do we use $P(n, (n_1, n_2, \ldots, n_k)) = \frac{n!}{n_1!n_2!\cdots n_k!}$ to calculate permutation?
6. **[combination]** $\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$. Explain the equation on the left. What's the relationship between combination and permutation? Does the order matter in combination?
7. **[probability properties]** Use Venn Diagram to show $\Pr(A \cup B) = \Pr(B) + \Pr(A) - \Pr(A \cap B)$. To generalize, what's $\Pr(A \cup B \cup C)$?
8. **[conditional probability]** $\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$. Explain the equation on the left. Can you make intuitive explanation for the joint probability $\Pr(A \cap B) = \Pr(A)\Pr(B|A)$?
9. **[law of total probability or marginalization]** $\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap A_i) = \sum_{i=1}^{n} \Pr(A_i)\Pr(B|A_i)$.
10. **[independence]** $\Pr(A \cap B) = \Pr(A)\Pr(B)$. What condition must be satisfied for the equation on the left to hold? Research on mutually independent and pair-wise independent.
11. **[random variable]** Random variable is a function that maps the sample space to a real value space.

    $$X : S \to \mathbb{R} \in R_X$$

    All the possible real values form a set called "range" R_X. Recall the pixel example in discussion 1, if I want a random variable to describe the grey level (use the simple average method: $I_{grey} = (R + G + B)/3$) of a single pixel, what's the range R_X?
12. **[equivalent events]** Choose any one pair of equivalent events from (1) and explain. Can you see why their probabilities are equal?
13. **[PMF / PDF]** For a random variable (remember it's a function) X, if its range R_X is finite or countably infinite, we call X **discrete** RV. On the other hand, if R_X is an interval or a collection of intervals, we call X **continuous** RV.

    For discrete RV, we use **probability mass function (PMF)** to describe the probability distribution of X; for continuous, we use **probability density function (PDF)** to describe the probability distribution of X.

    An ***important*** note is that PMF is a proper probability, but PDF is **NOT** a proper probability! Can you see why?
14. **[CDF]** $F(x) = \Pr(X \leq x)$. PDF for continuous random variable can be otained by $f(x) = \frac{dF(x)}{dx}$ is the derivative exists.
15. **[Expectation]** Expectation tells us the average (i.e., expected) value of some function $f(x)$ taking into account the distribution of x.

    $$E[f(x)] = \sum_x f(x)p(x)$$
    $$E[f(x)] = \int f(x)p(x)dx$$

    - If $f(x) = x$
      - $E[f(x)] = E(x) = \mu_x$, the "mean of x"
      - If we observe x many (infinite) times and average, we get μ_x

- If $f(x) = (x-\mu_x)^2$
  - $E[f(x)] = E\left[(x-\mu_x)^2\right] = \sigma_x^2$
  - $\sigma_x^2 = \mathrm{Var}(x)$ called "variance"; σ_x called "standard deviation"
  - If we observe each observation and μ_x, we get σ_x^2
  - Measure how likely x is going to be far away from the mean

16. [**properties**] For mean: $E(aX+b) = aE(X)+b$. For variance: $V(X) = E\left(X^2\right) - [E(X)]^2$, $V(aX+b) = a^2V(X)$.
17. [**Chebyshevs Inequality**] $\Pr(|X-\mu| > k\sigma) \le \frac{1}{k^2}$, given mean μ and variance σ for a random variable, and $\mathbf{k} > \mathbf{0}$. Another form is $\Pr(|X-\mu| \le k\sigma) \ge 1 - \frac{1}{k^2}$.
18. [**2-D random variable**] Let E be an experiment and S a sample space associated with E. Let X and Y be two functions each assigning a real number to each $s \in S$. We call (X,Y) a two-dimensional random variable (or random vector). Its **range** is $R_{X,Y} = \{(x,y) | x = X(s), y = Y(s), s \in S\}$. The definition can be extended to higher dimension, and they can be defined for both discrete and continuous random variable.
19. [**Joint probability function**]

    Discrete:

$$\begin{aligned} f_{X,Y}(x_i,y_j) &= \Pr(X=x_i, Y=y_j) \\ f_{X,Y}(x_i,y_j) &\ge 0 \text{ for all } (x_i,y_j) \in R_{X,Y} \\ \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f_{X,Y}(x_i,y_j) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \Pr(X=x_i, Y=y_j) = 1 \end{aligned} \tag{1}$$

    Continuous: change $\sum$ to $\int$.

20. [**Marginal distribution**] Recall the law of total probability (discrete case):

$$\begin{aligned} \Pr(A) &= \sum_n \Pr(A \cap B_n) \\ &= \sum_n \Pr(A|B_n)\Pr(B_n) \end{aligned} \tag{2}$$

    If (X,Y) is a 2-D discrete random variable, and its joint probability is $f_{X,Y}(x,y)$, the marginal distributions are:

$$\begin{aligned} f_X(x) &= \sum_y f_{X,Y}(x,y) \\ f_Y(y) &= \sum_x f_{X,Y}(x,y) \end{aligned} \tag{3}$$

21. [**Conditional distribution**]

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \text{ if } f_X(x) > 0 \tag{4}$$

22. [**Independence**]

$$\mathrm{P}(A\cap B) = \mathrm{P}(A)\mathrm{P}(B) \Longleftrightarrow \mathrm{P}(A) = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)} = \mathrm{P}(A|B) \tag{5}$$

23. [**Expectation**]

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y) f_{X,Y}(x,y), & \text{for Discrete RV}'\text{s} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy, & \text{for Continuous RV's} \end{cases} \tag{6}$$

    (a) A special case is that when $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$, the expectation is the **covariance** of (X,Y). $\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X\mu_Y$.
    (b) If X and Y are independent, $\mathrm{cov}(X,Y) = 0$. But $\mathrm{cov}(X,Y) = 0$ does **NOT** imply independence.
    (c) $\mathrm{Cov}(aX+b, cY+d) = ac\,\mathrm{Cov}(X,Y)$, $V(aX+bY) = a^2V(X) + b^2V(Y) + 2ab\,\mathrm{Cov}(X,Y)$

24. [**Correlation coefficient**]

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} \tag{7}$$

    (a) $-1 \le \rho_{X,Y} \le 1$
    (b) $\rho_{X,Y}$ measures the degree of **linear relationship** between X and Y.
    (c) If X and Y are independent, $\rho_{X,Y} = 0$. But $\rho_{X,Y} = 0$ does **NOT** imply independence.

Discrete distributions: Discrete uniform distribution, Bernoulli and Binomial distribution, Negative binomial distribution, Poisson distribution (and its approximation to Binomial distribution).

Continuous distributions: Continuous uniform distribution, Exponential distribution, Normal distribution (and its approximation to Binomial distribution).

25. **[Discrete uniform]** Equal probability for all discrete values. $f_X(x) = 1/k, \quad x = x_1, x_2, \cdots, x_k$, and 0 otherwise. Its mean and variance:

$$\mu = E(X) = \sum x f_X(x) = \sum_{i=1}^{k} x_i \frac{1}{k} = \frac{1}{k} \sum_{i=1}^{k} x_i \tag{8}$$

$$\sigma^2 = V(X) = \sum (x-\mu)^2 f_X(x) = \frac{1}{k} \sum_{i=1}^{k} (x_i - \mu)^2 \quad (= E\left(X^2\right) - \mu^2 = \frac{1}{k}\left(\sum_{i=1}^{k} x_i^2\right) - \mu^2) \tag{9}$$

26. **[Bernoulli and Binomial]** Random experiments with only **two possible outcomes** are defined as Bernoulli experiments. $f_X(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$, where $0 < p < 1$. We can also denoe as $X \sim Ber(p)$.

$$\mu = E(X) = p \tag{10}$$

$$\sigma^2 = V(X) = p(1-p) \tag{11}$$

If we take the Bernoulli trials for **n** times, with each trial being *independent*, and observe **x** times of success. We say the random variable X, where x is take from, is defined to have a binomial distribution: $\Pr(X = x) = f_X(x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x}$, for $x = 0, 1, \cdots, n$ and $0 < p < 1$. Also denote as $X \sim B(n, p)$. Notice that when $n = 1$, it becomes Bernoulli distribution.

$$\mu = E(X) = np \tag{12}$$

$$\sigma^2 = V(X) = np(1-p) \tag{13}$$

27. **[Negative binomial]** Let X be a random variable that represents the number of trials to produce k successes in a sequence of independent Bernoulli trials. X is said to follow a Negative Binomial distribution, namely $X \sim NB(k, p)$: $\Pr(X = x) = f_X(x) = \begin{pmatrix} x-1 \\ k-1 \end{pmatrix} p^k q^{x-k}$ for $x = k, k+1, k+2, \cdots$.

$$E(X) = \frac{k}{p} \tag{14}$$

$$\mathrm{Var}(X) = \frac{(1-p)k}{p^2} \tag{15}$$

Notice that the number of trials that are required to have the *first* success is known to follow a special case of negative binomial distribution called *geometric distribution*.

28. **[Poisson]** Experiments yielding numerical values of a random variable X, *the number of successes occurring during a given time interval or in a specified region*, are called Poisson experiments. And the number of successes X in a Poisson experiment is called a Poisson random variable, $X \sim Poisson(\lambda)$: $f_X(x) = \Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$, for $x = 0, 1, 2, 3, \cdots$.

$$E(X) = \lambda \tag{16}$$

$$V(X) = \lambda \tag{17}$$

Recall the Binomial distribution defined in (2), suppose that $n \to \infty$ and $p \to 0$ in such a way that $\lambda = np$ remains constant. We have X being approximated by a Poisson distribution:

$$\lim_{\substack{p \to 0 \\ n \to \infty}} \Pr(X = x) = \frac{e^{-np}(np)^x}{x!}$$

If $p \to 1$, we can still use the approximation by interchanging the definition of success and failure.

29. **[Continuous uniform]** A continuous random variable, which is uniformly distributed over the interval $[a, b]$, $-\infty < a < b < \infty$. $f_X(x) = \frac{1}{b-a}$, for $a \leq x \leq b$, and 0 otherwise.

$$E(X) = \frac{a+b}{2} \tag{18}$$

$$V(X) = \frac{1}{12}(b-a)^2 \tag{19}$$

30. **[Exponential]** A continuous random variable X assuming all nonnegative values is said to have an exponential distribution with parameter $\alpha > 0$ if its probability density function is given by $f_X(x) = \alpha e^{-\alpha x}$, for $x > 0$. Denote as $X \sim \text{Exp}(\alpha)$

$$E(X) = \frac{1}{\alpha} \tag{20}$$

$$V(X) = \frac{1}{\alpha^2} \tag{21}$$

No Memory Property of Exponential Distribution: for any two positive numbers s and t, $\Pr(X > s + t | X > s) = \Pr(X > t)$. *Meaning*: If X denotes the life length of a bulb, given that the bulb has lasted s time units, then the probability of it lasting for the next t time units is the same as the probability that it would last for the first t time units as brand new.

Another note is that the exponential distribution is frequently used as a model for the *distribution of times between the occurrence of successive events* such as customers arriving at a service facility or calls coming in to a switchboard.

31. **[Gaussian]** The PDF of Gaussian (normal) distribution is: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, $-\infty < x < \infty$, where $-\infty < \mu < \infty$ and $\sigma > 0$. Denote as $X \sim N\left(\mu, \sigma^2\right)$.

$$E(X) = \mu \tag{22}$$

$$V(X) = \sigma^2 \tag{23}$$

To obtain the standardized Gaussian, let $Z = \frac{(X-\mu)}{\sigma}$, and result in $Z \sim N(0, 1)$, $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z^2}{2}\right)$.

Statistical table: gives the values $\Phi(z)$ of a given z, where $\Phi(z)$ is the cumulative distribution function of a standardized Normal random variable Z.

$$\begin{aligned} \Phi(z) &= \Pr(Z \leq z) \\ 1 - \Phi(z) &= \Pr(Z > z) \end{aligned} \tag{24}$$

Recall the Binomial distribution defined in (2), suppose that $n \to \infty$ and $p \to \frac{1}{2}$ (or even when n is small and p is not extremely close to 0 or 1), we have X being approximated by a Gaussian distribution with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$:

$$Z = \frac{X - np}{\sqrt{npq}} \text{ is approximately } \sim N(0, 1)$$

32. **[Population, sample]** A ***population*** is a set of similar items or events which is of interest for some question or experiment, and a ***sample*** is any subset of population. Every outcome or observation can be recorded as a *numerical* or a *categorical* value. Population may be finite or infinite.

33. **[Random sampling]** ***Simple random sample*** of n observations is a sample such that every subset of n observations of the population has the same probability of being selected.

    (a) When we sample from a finite population, we can sample with/without replacement. This corresponds to the counting problems.

    (b) When we sample from an infinite population, if we assume that all random variables have the ***same*** distribution and are ***independent***, we say that the sample is random.

34. **[Sampling distribution]** The main purpose of sampling is to estimate some ***unknown population parameters***, so that we can make some ***inference*** regarding the true population. A value computed from a sample is called a ***statistic***, and it varies (why?). Hence a statistic should be a random variable. The *probability distribution of a statistic* is called a ***sampling distribution***.

Sample mean defined by the statistic:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{25}$$

Theorem (see an example): For random samples of size n taken from an *infinite* population or from a *finite population with replacement* having population mean μ and population standard deviation σ, the **sampling distribution of the sample mean** has:

$$\mu_{\overline{X}} = \mu_X \quad \text{and} \quad \sigma^2_{\overline{X}} = \frac{\sigma^2_X}{n} \tag{26}$$

Law of large number (LLN): Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population having any distribution with mean μ and *finite* population variance σ^2. Then for any $\epsilon \in \mathcal{R}$

$$P(|\overline{X} - \mu| > \epsilon) \to 0 \text{ as } n \to \infty \tag{27}$$

Central limit theorem (CLT): Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population having any distribution with mean μ and *finite* population variance σ^2. If n is sufficiently large, the sampling distribution of the sample mean $\overline{X}$ is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \text{ approx } \sim N(0,1) \tag{28}$$

Theorem: if $X_i, i = 1, 2, \cdots, n$ are $N\left(\mu, \sigma^2\right)$, the sample mean $\overline{X}$ is $N\left(\mu, \frac{\sigma^2}{n}\right)$ regardless of the sample size n.

What about the **sampling distribution of the difference of two sample means**? If independent samples of size $n_1 (\geq 30)$ and $n_2 (\geq 30)$ are drawn from two large or infinite populations, discrete or continuous, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. The sampling distribution of the difference of means, $\overline{X}_1$ and $\overline{X}_2$, is approximately normally distributed with mean and standard deviation given by $\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$ and $\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$:

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \text{ approx } \sim N(0,1) \tag{29}$$

35. **[Chi-square distribution]** Y is a chi-square distribution with n *degrees of freedom* if

$$f_Y(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}, \qquad \text{for } y > 0, \text{and } 0 \text{ otherwise} \tag{30}$$

It is denoted as $\chi^2(n)$, and n is a positive integer and $\Gamma(\cdot)$ is the gamma function: $\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx = (n-1)!$.

$E(Y) = n$, $V(Y) = 2n$; if $X \sim N(0,1)$, then $X^2 \sim \chi^2(1)$; let $X_1, X_2, \cdots, X_n$ be a random sample from a normal population with mean μ and variance σ^2, $Y = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$.

36. **[Estimation based on Normal Distribution]** Given the observed data $x_1, x_2, \cdots, x_n$, we want to estimate the parameter θ which controls the distribution $f_X(x|\theta)$. A statistic is a function of the random variable which *does not depend on any unknown parameters*. The statistic that one uses to obtain a point estimate is call an **estimator**. Interval estimation is to define two statistics and use their interval to estimate the parameters.

Unbiased estimator: $E(\widehat{\Theta}) = \theta$.

Confidence interval for interval estimation[1]. For the given error margin, the sample size is given by $n \geq \left(Z_{\alpha/2}\frac{\sigma}{e}\right)^2$.

Confidence interval for the mean in 1) known variance case; 2) unknown variance case.

Confidence interval for the difference between two means. $\overline{X}_1 - \overline{X}_2$ is a point estimator of $\mu_1 - \mu_2$. Also two cases, known variances and unknown variances.

Confidence interval for the difference between two means for paired data (dependent data).

Confidence interval for a variance.

Confidence interval for the ratio of two variances with unknown means.

37. **[Hypotheses testing based on Normal Distribution]**

- Often, hypothesis is stated in a form that hopefully will be rejected, denoted as H_0 (Null hypothesis); its opposite (the one we need to accept due to insufficient data for concluding false) is denoted as H_1(Alternative hypothesis).
- Two tailed test and one tailed test.
- Type I and Type II error.
- Acceptance and rejection regions, critical value.
- Hypothesis testing on mean with known/unknown variance.
  - Two-sided
  - One-sided
- Hypothesis testing on difference between two means.
- Hypothesis testing on variance.
- Hypothesis testing on ratio variance.

[1] p21