# Review 1 (on Chapter 5(B) —)

March 17, 2011

## 1 Some important random variables

- $X$ takes n values, $x_1, ..., x_n$ with probability distribution function

$$P(X = x_k) = p_k$$

e.g. toss a even die, n = 6 and P(X=1) = ... = P(X=6) = 1/6

- Binomial distribution $X \sim B(n, p)$
- Poisson distribution $X \sim Poi(\lambda)$.

- Uniform distribution on [a, b], the probability density function

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

- Normal distribution $N(\mu, \sigma^2)$

- For continuous variables, the pdf, $f(x)$, and CDF, $F(x)$, are defined on the whole real space.

- We did not talk in details about $F(x)$, you only need to know it is an nondecreasing nonnegative function with $F(-\infty) = 0$ and $F(\infty) = 1$.

## 2 Expectation and Variance of a random variable

- Expectation, usually denoted by $\mu = EX$

  1. discrete $EX = \sum_{x_i} x_i p_i$
  2. continuous $EX = \int_{-\infty}^{\infty} x f(x) dx$

- Expectation of function of random variable

  1. discrete $E[g(X)] = \sum_{x_i} g(x_i) p_i$
  2. continuous $E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$

- Variance, usually denoted by $\sigma^2 = E\{(X-\mu)^2\} = EX^2 - (EX)^2$

  1. discrete $Var(X) = \sum_{x_i}(x_i - \mu)^2 p_i$
  2. continuous $Var(X) = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$

- You need to know the expectation and variance for the distributions listed above and some simple calculations

**Example** The manager of a bakery knows that the number of chocolate cakes he can sell on any given day is a random variable with probability mass function

$$p(x) = 1/6 \quad \text{for } x = 0, 1, 2, 3, 4 \text{ and } 5.$$

He also knows that there is a profit of $1.00 on each cake that he sells and a loss (due to spoilage) of $0.40 on each cake that he does not sell. Assuming that each cake can be sold only on the day it is made, find the baker's expected profit for a day when he bakes 5 chocolate cakes.

**Example** If a contractor's profit on a construction job is a random variable with probability density function

$$f(x) = \begin{cases} \frac{1}{18}(x+1), & \text{if } -1 < x < 5 \\ 0, & \text{otherwise} \end{cases}$$

where the units are \$1000, what is her expected profit?

5.47 From experience Mr. Harris has found that the low bid on a construction job can be regarded as a random variable having the uniform density

$$f(x) = \begin{cases} \dfrac{3}{4C} & \text{for } \dfrac{2C}{3} < x < 2C \\ 0 & \text{elsewhere} \end{cases}$$

where $C$ is his own estimate of the cost of the job. What percentage should Mr. Harris add to his cost estimate when submitting bids to maximize his expected profit?

**Example**: In an urn, there are six cards numbered 1, 2, ..., 6 respectively. Randomly draw 2 cards without replacement. Let $X_1$ and $X_2$ be the numbers on the first card and second card respectively. Find their joint distribution.

**Example** Let X be the total number of items produced in a day's work at a factory, and let Y be the number of defective items produced. Suppose that the probability mass function for (X; Y) is given by Table below.

| | | $Y$ | | | | |
|---|---|---|---|---|---|---|
| $p_{X,Y}(x,y)$ | | 1 | 2 | 3 | 4 | 5 |
| | 1 | .05 | 0 | 0 | 0 | 0 |
| | 2 | .15 | .10 | 0 | 0 | 0 |
| $X$ | 3 | .05 | .05 | .10 | 0 | 0 |
| | 4 | .05 | .025 | .025 | 0 | 0 |
| | 5 | .10 | .10 | .10 | .10 | 0 |

(a) Find the marginal distribution of $Y$, i.e. $P(Y = 1), ..., P(Y = 5)$

(b) Find $E(X + Y)$ and $E(XY)$

5.72 Two random variables are independent and each has a binomial distribution with success probability 0.4 and 2 trials.

(a) Find the joint probability distribution.

(b) Find the probability that the second random variable is greater than the first.

## 4 More about Expectation and Variance

- For any random variables $X_1, ..., X_n$ and constants $a_0, a_1, ..., a_n$, we have

$$E(a_0 + a_1X_1 + ... + a_nX_n) = a_0 + a_1E(X_1) + ... + a_nE(X_n)$$

- if random variables $X_1, ..., X_n$ are independent, then

$$Var(a_0 + a_1X_1 + ... + a_nX_n) = a_1^2Var(X_1) + ... + a_n^2Var(X_n)$$

- For random sample $X_1, ..., X_n$ from a population with mean $\mu$ and variance $\sigma^2$, let $\bar{X}$ be the sample mean and $S^2$ be the sample variance

(a) $EX_i = \mu, Var(X_i) = \sigma^2$, (b) $E\bar{X} = \mu$
(c) $Var(\bar{X}) = \sigma^2/n$, (d) $E(S^2) = \sigma^2$

**Example** The foreman of a bottling plant has observed that the amount of soda in each "32-ounce" bottle is actually a normally distributed random variable, with a mean of 32.2 ounces and a standard deviation of .3 ounce. If a customer buys a carton of four bottles, what is the probability that the mean amount of the four bottles will be greater than 32 ounces?

| Population | $\sigma$ | n | Statistic and its distr. |
|---|---|---|---|
| Normal | known | small or large. | $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ |
| | unknown | small | $t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ |
| any | known | large | $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ |
| | unknown | large | $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$ |

**Example** The number traffic accidents in a city follows Poisson distribution with $\lambda = 5$. What is the probability for the total number of traffic accidents in a year to exceed 2000?

# 6 Estimation

A lumber company must estimate the mean diameter of trees to determine whether or not there is sufficient lumber to harvest an area of forest. They need to estimate this to within 1 inch at a confidence level of 99%. The tree diameters are normally distributed with a standard deviation of 6 inches. How many trees need to be sampled?

Things we know: Normal distribution population with $\sigma = 6$, confidence 99%, therefore $\alpha = .01$ and $z_{\alpha/2} = 2.575$, $E = 1$. Thus

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = 239$$

That is, we will need to sample at least 239 trees.

A random sample of size n $= 100$ is taken from a population with $\sigma = 5.1$. Given that the sample mean is $\bar{x} = 21.6$. construct a 95% confidence interval for the population mean $\mu$

Substituting the given values o f $n, \bar{x}, \sigma$, and $z_{0.025} = 1.96$ into the confidence interval formula $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ since $n > 30$, we get

$$21.6 - 1.96 \times \frac{5.1}{\sqrt{100}} < \mu < 21.6 + 1.96 \times \frac{5.1}{\sqrt{100}}$$

or $20.6 < \mu < 22.6$. Of course, either the interval from 20.6 to 22.6 contains the population mean $\mu$ or it does not, but we are 95% confident that it does.

With reference to the nanopillar height data below, 245, 333, 366, 323, 266, 391, 315, 355, 308, 276, 296, 304, 276, 336, 289, 234, 309, 284, 310, 338, 297, 314, 315, 305, 290, 300, 292, 311, 346, 337, 303, 265, 278, 276, 364, 390, 298, 290, 308, 221, 253, 292, 305, 330, 272, 312, 373, 271, 274, 343, construct a 99% confidence interval for the population mean of all nanopillars.

By using $\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$ with n= 50, $\bar{x} = 305.58nm$, and $s^2 = 1,366.86$ (hence, $s = 36.97nm$) and $z_{\alpha/2} = z_{0.005} = 2.575$, we get

$$292.12 < \mu < 319.04.$$

We are 99% confident that the interval from 292.12 nm to 319.04 nm contains the true mean nanopillar height.

The mean weight loss of n = 16 grinding balls after a certain length of time in mill slurry is 3.42 grams with a standard deviation of 0.68 gram. Construct a 99% confidence interval for the true mean weight loss of such grinding balls under the stated conditions.

By using $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$ with $n = 16, \bar{x} = 3.42, s = 0.68,$ and to $t_{.005} = 2.947$ for n - 1 = 15 degrees of freedom, we get

$$3.42 - 2.947 \times \frac{0.68}{4} < \mu < 3.42 + 2.947 \times \frac{0.68}{4}$$

or $2.92 < \mu < 3.92$. We are 99% confident that the interval from 2.92 grams to 3.92 grams contains the mean weight loss.