Statistics ST2334 Topic 4: Probability Distributions (part c)

2011/2012 Semester 2

Cumulative Distribution Function

- The cumulative distribution function (cdf) for a random variable X is
$$F(x) = P(X \leq x)$$
- For discrete X, its probability mass function $f(x)$, as well as the cdf $F(x) = \sum_{y \leq x} f(y)$ characterizes its distribution.
- The cdf extends naturally to other types of random variables. For this reason, it is sometimes also known simply as the distribution function.

Example: Geometric Distribution cdf

Let $X \sim Geo(p)$. Then its cdf is

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \sum_{k=1}^{x} P(X = k) \\ &= \sum_{k=1}^{x} (1-p)^{k-1} p \\ &= p \cdot \frac{1-(1-p)^x}{1-(1-p)} \\ &= 1-(1-p)^x \end{aligned}$$

Independent Random Variables

Let X, Y be independent random variables. Then,

- $E(XY) = E(X)E(Y)$

$$\begin{aligned} E(XY) &= \sum_x \sum_y xyP(X=x, Y=y) \\ &= \sum_x \sum_y xyP(X=x)P(Y=y) \\ &= \sum_x xP(X=x) \sum_y yP(Y=y) = E(X)E(Y) \end{aligned}$$

- $Var(X+Y) = Var(X) + Var(Y)$

Example: Verify Binomial Expectation and Variance

- Let $X_1, X_2, \ldots X_n$ be independent Bernoulli random variables with probability of success p.
- Let $Y = \sum_{i=1}^{n} X_i$. Then we know $Y \sim Bin(n, p)$

$$E(X_i) = p \text{ and } E(X_i^2) = p \Rightarrow Var(X_i) = p - p^2 = p(1-p)$$

- Hence,

$$E(Y) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = nE(X_i) = np$$

$$Var(Y) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) = nVar(X_i) = np(1-p)$$

Sampling from Population

- For a population with mean μ and variance σ^2, X is an random sample/observation from the population.
- Then
$$E[X] = \mu, \qquad Var[X] = \sigma^2$$
- and
$$\text{the distribution of } X = \text{ the distribution of the population}$$
- For large populations, taking a sample of size n is like observing n random variables $X_1, X_2, \ldots, X_n$, independent and identically distributed (iid) copies of X.

Sample Mean

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ represent the sample mean. Then,

$$E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$$

$$Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n}\sigma^2$$

"Sample Variance"

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] &= E\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right] \\
&= \sum_{i=1}^{n} E\left[X_i^2\right] - nE\left[\bar{X}^2\right] \\
&= nE\left[X_i^2\right] - nE\left[\bar{X}^2\right] \\
&= n(Var[X_i] + (E[X_i])^2) - n(Var\left[\bar{X}\right] + (E\left[\bar{X}\right])^2) \\
&= n(\sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2) \\
&= (n-1)\sigma^2
\end{aligned}
$$

Therefore,

$$
E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \sigma^2
$$

Example: A Fair Game?

- Suppose we are offered to play a game of chance under these conditions:
  - It costs us to play \$1.5 and the awarded prices are \$1, \$2, \$3.
  - The probabilities of winning each price are $\{0.6, 0.3, 0.1\}$, respectively.
  - Should we play the game?
- Let \$$X$=awarded price. Then $X=\{1, 2, 3\}$.

x	1	2	3
P(X=x)	0.6	0.3	0.1

$$E[X] = 1 \cdot 0.6 + 2 \cdot 0.3 + 3 \cdot 0.1 = 1.5$$

- The cost and the expected prize is the same. We call this a fair game.

Example: Find Standard Deviation for the above game

The variance for this game is computed by

$$\begin{aligned} Var[X] &= (x_1 - 1.5)^2 P(X = x_1) \\ &\quad + (x_2 - 1.5)^2 P(X = x_2) \\ &\quad + (x_3 - 1.5)^2 P(X = x_3) \\ &= 0.45. \end{aligned}$$

Thus, the standard deviation of X is 0.671.

- You may find that the alternative formula for $Var[X]$ is easier to compute.

Example: Baby Gender

- Suppose we conduct an experiment involving young couple planning to have children.
- The couples are interested in the number of girls they will have, and each couple agrees to have children until one of the following 2 stopping criteria is met:
  1. the couple has at least one child of each gender,
  2. the couple has at most 3 children,
- So, $\Omega = \{\text{BBB, BG; GB; BBG, GGB, GGG}\}$
- Let X, Y be the number of Girls and Children respectively.

x	0	1	2	3
P(X=x)	1/8	5/8	1/8	1/8

y	2	3
P(Y=y)	1/2	1/2

$$E[X] = 0 + 5/8 + 2/8 + 3/8 = 1.25.$$

and

$$E[Y] = 2.5.$$

Example: Baby Gender

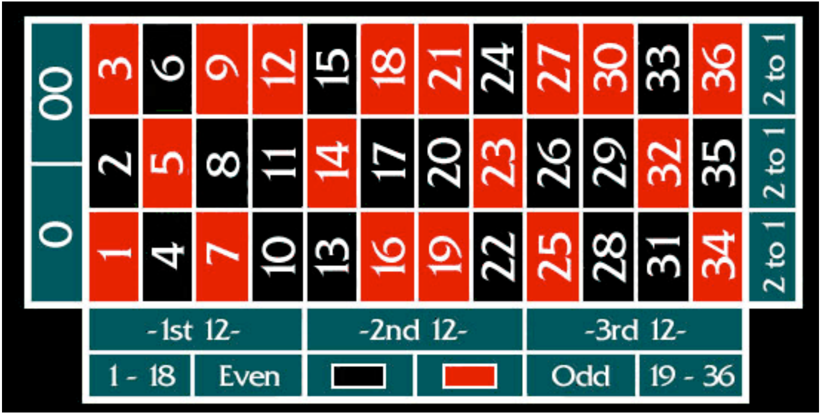
- The above stopping criteria gave us a (long run) average of 1.25 girls and 1.25 boys.
- Can we have a strategy that changes this (no abortions allowed)?
- Consider the strategy of keep giving birth until the first girl. Do we get more girls than boys this way?

Roulette

Roulette

- The wheel spins one way, and the ball spins the opposite direction eventually dropping into one of 38 possible pockets.
- The 38 outcomes numbered $0, 00, 1, \ldots, 36$ are equally likely.
- 0 and 00 are green and among the other 36, 18 of them are red and 18 of them black.
- Each spin is independent of other spins.

Roulette



- Many different ways to bet, each with different pay out.

Roulette

- Betting on Black pays 1 to 1. i.e. If you bet \$1 on Black you win \$1 if the ball lands on a black number, and you lose the \$1 otherwise.
  (a) What is the Expected Winnings if you make one \$1 bet on Black?
  (b) What is the Expected Winnings if you make one \$5 bet on Black?
  (c) What is the Expected Winnings if you make five \$1 bet on Black (over five spins)?
  (d) Variances? Which bet do you prefer?

Roulette (continued)

- You can also place bets on the corners of the numbers. For example you can bet on 1,2,4,5 by placing money at the intersection of the four boxes on the table. The bet pays 8 to 1.
  - (e) What is the Expected Winnings if you make one \$1 bet on "1,2,4,5"?
  - (f) Variance? Compared to (a), which bet do you prefer?