MATH 3338 Probability

Wenjiang Fu
wfu2@central.uh.edu

Lecture 5 - Chapter 5. Important Distributions and Densities

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

Important Distributions

- Describe discrete prob distributions and continuous densities.
- **Discrete Uniform Distribution** RV X is uniformly distributed over n elements in the sample space Ω consisted of n elements, if the distribution function $m(\omega) = 1/n$ for each element $\omega \in \Omega$.
- **Binomial Distribution** A Bernoulli trial is repeated n times independently, each time it has the same prob p to take the value 1 and $1 - p$ to take 0. Then the sum of these n indep trials has the Binomial distribution with the prob of being exactly k with $0 \leq k \leq n$

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Important Distributions

- **Geometric Distribution**
  Consider a Bernoulli trials process continued for an infinite number of trials. Let T be the number of trials up to and including the first success. Then,

  $$P(T = 1) = p; P(T = 2) = qp; P(T = 3) = q^2 p; ...$$

  In general, $P(T = n) = q^{n-1} p$. Fact: $p + qp + q^2 p + ... + = 1$.

- In general, if $0 < p < 1$, and $q = 1 - p$, then we say that rv T has a geometric distribution if

  $$P(T = j) = q^{j-1} p,$$

  for $j = 1, 2, 3...$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

- **Geometric Distribution**
  Example 5.1 Consider a queuing or waiting line of customers waiting for service. It is often assumed that, in each small time unit, either 0 or 1 new customer arrives. The prob a customer arrives is p and none arrives is $q = 1 - p$. Then the time T until the next arrival has a geometric distr. It is natural to ask for the prob that no one arrives in the next k time units, or $P(T > k)$.

$$P(T > k) = \sum_{j=k+1}^{\infty} q^{j-1} p = q^k (p + qp + ...) = q^k.$$

  This is exactly the prob of having exactly k failures (no one arrives), thus the prob q^k.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Important Distributions

- **Geometric Distribution**
  Consider a different senario. It's often assumed the length of time required to service a customer also has a geometric distr. but with a different p. Now consider the conditional prob

$$P(T > r + s | T > r) = \frac{P(T > r + s)}{P(T > r)} = \frac{q^{r+s}}{q^r} = q^s.$$

  It is implied that the customer's service takes s more time units is indep of the length of time r that (s)he has already been serviced. This is the so-called “memoryless" feature, similar to the one of the exponential distribution, but this one is for discrete RV.

- **Poisson Distribution**
  One of the three most important discrete distr.
  Consider the event of traffic accidents. Suppose in a specific district, such as a segment of free way (I-10). We count the number of accidents on one day. Suppose the traffic remains constant (non-rush hour). Three basic assumptions.
  1) Rare event. Within each small unit time, there is either 1 accident, or no accident. $P(> 1 accidents) = 0$.
  2) Stationary. Within the same length of time window (eg. hour), the intensity of accidents remains the same at different time of the day.
  3) Independence. The outcome of having accidents in one time interval does not affect the likelihood of having accidents at another time interval that has no time overlapping.

- **Poisson Distribution**
  Assume the intensity of accidents per unit time is λ. Then the prob of having k accidents in a time interval of length t is

$$P(X=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t},\ \ k=0,1,2,...$$

Often, the time duration t can be absorbed into the intensity, or a large unit $t=1$. It is simplified as

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},\ \ k=0,1,2,...$$

Note that there is no upper limit of the value k that X may take. In theory, the number can go up as high as possible, but the prob becomes very small, no chance to occur.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Important Distributions

- **Negative Binomial Distribution**
  Recall that we learned Geometric distribution. In tossing a coin with prob p to get a head. We ask the prob of having the k-th toss to get the first head. The prob is a geometric distribution

$$P(X = k) = (1 - p)^{k-1} p$$

  Now consider the prob of getting the k-th head at the x-th toss. Notice that there are $k - 1$ heads in the previous $x - 1$ tosses. So for each set of tossing, the probability of having a total of k heads out of x tosses is $q^{x-k} p^k$. But since there are $k - 1$ heads out of the first $x - 1$ tosses, there are C_{k-1}^{x-1} sets of different arrangements. Hence we have the Negative Binomial distribution

$$P(X = x) = \binom{x-1}{k-1} p^k q^{x-k}$$

  The negative binomial RV X is the sum of k indep geometrically distributed RV Y_i with prob p, i.e. $X = \sum_{i=1}^{k} Y_i$.

- **Hypergeometric Distribution**
  Suppose there are a total of N balls, with k red balls and $N-k$ blue balls. We choose n balls without replacement. Then the prob of having exact x red balls is

$$P(X=x)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

Remark 1.This prob distribution uses the combination number and the number of ways to complete a task in 2 steps.
Remark 2. Let $k, N \to \infty$, and $k/n \to p$ and $0<p<1$. Then the hypergeometric prob converges to the Binomial prob with prob p, i.e $Bin(n,p)$.

The definition of hypergeometric distribution has many applications (e.g Defect products). The distribution can be also extended to more than two types of objects.

Important Densities

- **Continuous Uniform Distribution**
  Recall that we have defined continuous uniform distribution over an interval $[a, b]$. The density $f(x)$ is defined as

$$f(x) = \frac{1}{b-a} I_{[a,b]}$$

where $I_{[a,b]} = 1$ for $x \in [a, b]$, and 0 otherwise, is an indicator function of the interval $[a, b]$.

Exponential and Gamma Densities The exponential density function is defined by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } 0 \le x < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Where $\lambda > 0$.

The CDF function $F(x)$

$$F(x) = \int_{-\infty}^{x} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}.$$

- **Exponential Distribution**
  The exponential distribution has the memoryless property

$$P(T > r + s | T > r) = \frac{P(T > r + s)}{P(T > r)} = \frac{1 - F(r + s)}{1 - F(r)} = 1 - F(s)$$

$$= P(T > s) = e^{-\lambda s}$$

- **Functions of a Random Variable**

Theorem 5.1 Let X be a continuous RV, and $\phi(x)$ is a strictly increasing function on the range of X . Define $Y = \phi(X)$. Suppose that X and Y have CDFs F_X, F_Y. then we have

$$F_Y(y) = F_X(\phi^{-1}(y)).$$

If $\phi(x)$ is strictly decreasing on the range of X, then

$$F_Y(y) = 1 - F_X(\phi^{-1}(y))$$

Proof Assume ϕ is strictly increasing, then

$$F_Y(y) = P(Y \leq y) = P(\phi(X) \leq y) = P(X \leq \phi^{-1}(y)) = F_X(\phi^{-1}(y))$$

Assuming ϕ is strictly decreasing.

$$F_Y(y) = P(Y \leq y) = P(\phi(X) \leq y) = P(X \geq \phi^{-1}(y))$$
$$= 1 - F_X(\phi^{-1}(y))$$

- **Functions of a Random Variable**
  Corollary 5.1 Let X be a continuous RV, and $\phi(x)$ is a strictly increasing function on the range of X. Define $Y = \phi(X)$. Suppose the density functions of X and Y are f_x and f_Y, respectively. Then

$$f_Y(y) = f_X(\phi^{-1}(y))\frac{d}{dy}\phi^{-1}(y).$$

  If $\phi(x)$ is strictly decreasing on the range of X, then

$$f_Y(y) = -f_X(\phi^{-1}(y))\frac{d}{dy}\phi^{-1}(y)$$

  Proof Take derivative on both sides of equation in Theorem 5.1.

Important Densities

- **Functions of a Random Variable**
  Example RV X has a density function $f(x) = 3x^2$ over the interval $[0, 1]$. Let $Y = X^3$. Find the density of Y.
  Solution $\forall y \in [0, 1]$, $Y = X^3$. Then $X = Y^{1/3}$.

$$f_Y(y) = f_X(y^{1/3})\frac{d}{dy}y^{1/3} = 3(y^{1/3})^2\frac{1}{3}y^{-2/3} = 1$$

  Hence Y has a uniform distribution over $[0, 1]$.

- **Simulation** Theorem 5.1 informs that if we take a special function of a RV X, and X has a CDF $F(x)$. Then $Y = F^{-1}(X)$ follows a uniform distribution over $[0, 1]$.
  The implication of this fact is, if we want to generate random variates X_i of a distribution with CDF F, we need to generate random variates Y_i following $Unif[0, 1]$, and then take $X_i = F^{-1}(Y_i)$. We know that X_i follows a distribution with CDF $F(x)$.

Important Densities

- **Functions of a Random Variable**
  Corollary 5.2 If $F(y)$ is a given CDF that is strictly increasing when $0 < F(y) < 1$ and if U is a random variable with uniform distribution on $[0, 1]$, then $Y = F^{-1}(U)$ has the CDF $F(y)$.

- **Normal Density** The normal density function with parameters μ and σ is defined as follows.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$$

  The parameter μ is the center or mean of the distribution, σ is the standard deviation. The CDF is

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma}} e^{-(u-\mu)^2/2\sigma^2} du$$

- **Functions of a Random Variable**
  Normal Distribution and Standard Normal Distribution
  Let Z be a RV following a normal distribution with parameter $\mu = 0, \sigma = 1$. Let $X = \sigma Z + \mu$, then X follows a normal distribution with parameters mean μ and standard deviation σ, ie. $N(\mu, \sigma^2)$.

- Similarly, suppose a RV X following a normal distribution with parameters mean μ and standard deviation σ, then we can always standardize X to make a RV following a standard normal distribution $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

- Since standard normal distribution plays a major role in calculating probability, we name a function $\Phi(z) = F(z)$, CDF of $N(0, 1)$. $\Phi(0) = 0.5$, $\Phi(1) = 0.841$, $\Phi(1.96) = 0.975$, $\Phi(2) = .9772$ and $\Phi(3) = .99865$. This means that within 1 and 2 standard deviation, $P(|Z| \leq 1) = .682$, $P(|Z| \leq 2) = .954$, $P(|Z| \leq 3) = .997$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

- **Functions of a Random Variable**
  Normal Distribution and Standard Normal Distribution
  With the standardization, $X = \sigma Z + \mu$,
  $P(|X - \mu|/\sigma \leq s) = P(|Z| \leq s)$. Hence
  $P(|X - \mu|/\sigma \leq 1) = .682$, $P(|X - \mu|/\sigma \leq 2) = .954$, and
  $P(|X - \mu|/\sigma \leq 3) = .997$

- **A Rule of Thumb.** For any distribution, except for strange distributions, the prob mass is .68 within 1 sd, .95 within 2 sd, and 99 within 3 sd.

  Example 5.8 Suppose $X \sim N(10, 3^2)$, Find $P(4 < X < 16)$.
  $P(4 < X < 16) = P((4 - 10)/3 < (X - \mu)/\sigma < (16 - 10)/3) =$
  $P(-2 < Z < 2) = .954$

- **Functions of a Random Variable**
  Chi-squared Distribution
  For a given RV $Z \sim N(0,1)$, we define RV $X = Z^2$. Then X is said to follow a Chi-squared distribution with 1 degree of freedom, χ_1^2.
  For a given sequence of indep RV $Z_1, ..., Z_n \sim N(0,1)$, we define a RV $Y = Z_1^2 + ... + Z_n^2$. Then Y is said to follow a Chi-squared distribution with n degrees of freedom, χ_n^2.