2019 Statistics Graduate Bootcamp
University of California, Irvine

TA: Derenik Haghverdian
Department of Statistics

Updated - September 12, 2019

1 Binomial

An experiment (a random physical process) is a *Binomial experiment* if it satisfies the following conditions:

- There are a fixed number of trials.
- Each trial has two possible outcomes, Success and Failure.
- The probability of success p on any trial is constant.
- The outcome of each trial is independent of the other trials.

If X is the number successes from a binomial experiment with n trials and p probability of success, then X is a random variable that follows the $Binomial(n, p)$ distribution with pmf

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \cdots, n.$$

1. Research on the internet or use your imagination to come up with a few examples of random processes that follow a binomial distribution.
2. Verify the pmf of the binomial distribution is a valid pmf.
3. Plot the pdf for a few different choices of parameter values.
4. Find the expected value of $X \sim Binom(n, p)$.
5. Find the mgf $M_X(t) = \mathbb{E}e^{tX}$.
6. Find $P(X \leq 2)$ for $X \sim Binom(8, 0.2)$. Sketch the pmf for this random variable and indicate the probability you calculated on the sketch.
7. A new cancer drug kills 99% of cancer cells in a laboratory setting. Suppose 2000 cancer cells are in a petri dish when the drug is applied, and let X = number of surviving cells.
   (a) Does this experiment follow a binomial distribution?
   (b) Write down but do not evaluate an expression for the probability that fewer than 100 cells survive using a binomial distribution model.

(c) The Poisson approximation to the Binomial states that $P(X = x) \approx P(Y = x)$ for $X \sim Binom(n, p), Y \sim Pois(\lambda = np)$, for n large and p small. Evaluate the probability from (b) using a Poisson approximation.

(d) The Binomial(n, p) distiribution can also be approximated by the normal distribution $N(\mu = np, \sigma^2 = np(1 - p))$ for n large. Approximate the probability from (b) using the correct normal distribution. Does the normal or Poisson approximation work better in this scenario?

(e) Use `R` to plot the Poisson and Normal approximations to the pmf for a Binomial$(20, p)$ distribution for $p = 0.1$ and $p = 0.5$.

2 Poisson

Many counting processes in the world approximately follow the Poisson distribution. For a random process X to be well-modelled a Poisson it should satisfy the following assumptions:

- X is the number of times an an event occurs in an interval and X can take values $0, 1, 2, \cdots$.
- Events occur independently.
- The rate of occurrence of events is constant.
- Two events cannot occur simultaneously.
- The probability of an event in an interval of time is proportional to the length of that interval.

The $Poisson(\lambda)$ distribution has pmf:

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \cdots .$$

1. Research on the internet or use your imagination to come up with a few examples of random processes that follow a Poisson distribution.
2. Plot the pdf for a few different choices of parameter values.
3. Use the Taylor series for e^λ to show that the Poisson pmf is actually a pmf.
4. For $X \sim Pois(\lambda)$, find $\mathbb{E}X$.
5. Find the moment generating function $M_X(t) = \mathbb{E}e^{tX}$. Use the mgf to find the variance of the Poisson distribution.
6. Let $X_1 \sim Pois(\lambda_1)$ and $X_2 \sim Pois(\lambda_2)$ where X_1 is independent of X_2. What is the distribution of $X_1 + X_2$?
7. An insect lays a large number of eggs, each surviving with probability p. Let X be the number of survivng eggs, and let N be the total number of eggs laid. Suppose $N \sim Pois(\lambda)$ and $X|N \sim Binom(N, p)$. Find $P(X = x)$. [This is Example 4.4.1 and 4.4.2 in Casella & Berger.]

3 Normal

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}.$$

1. Research on the internet or use your imagination to come up with a few examples of random processes that follow a normal distribution.
2. Plot the pdf for a few different choices of parameter values.
3. Show the pdf of $N(0,1)$ integrates to 1 over $\mathbb{R}$. You must use a polar coordinate transformation on the double integral
$$\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{1}{2\pi\sigma^2} \exp\{-\frac{1}{2\sigma^2}\left[(x-\mu)^2 + (y-\mu)^2\right]\}.$$
4. Estimate $P(5 > X > 3)$ for $N(1,4)$ using the Empirical Rule. (The Empirical Rule says that the area under a normal curve within 1 σ (standard deviation) of the mean is approximately 0.68; within 2σ of the mean is approximately 0.95; within 3σ of the mean is approximately 0.997.)
5. Find the mgf.
6. For $X \sim N(\mu,\sigma^2)$, verify $\mathbb{E}X = \mu$ and $VarX = \sigma^2$.
7. The Central Limit Theorem says that large sums of independent identically distributed random variables are approximately normal. A commonly used result states that for large n, a binomial(n,p) can be approximated by $N(np, np(1-p))$.
8. Let $X_1 \sim N(\mu_1,\sigma_1^2)$ and $X_2 \sim N(\mu_2,\sigma_2^2)$ and assume X_1 and X_2 are independent. Find the distribution of $X_1 + X_2$ by finding the mgf.

4 Exponential

$$f(x;\theta) = \theta \exp\{-\theta x\}, x > 0, \theta > 0.$$

1. Research on the internet or use your imagination to come up with a few examples of random processes that follow an exponential distribution.
2. Plot the pdf for a few different choices of parameter values.
3. Show the exponential pdf is a valid pdf.
4. Find the cdf $F(x)$ of $X \sim Exp(\theta)$.
5. For $Y \sim Exp(3)$, what is $P(1 < Y < 5)$? Illustrate this probability graphically.
6. Find $\mathbb{E}X$.
7. Find the mgf $M_X(t) = \mathbb{E}e^{tX}$.
8. A distribution with is said to be *memoryless* if

$$P(T > t + s | T > t) = P(T > s).$$

   Show that the exponential distribution is memoryless.
9. Let X_1, X_2 be independent and identically distributed $Exp(\theta)$. Show that $X_1 + X_2 \sim Gamma(2, \theta)$ by finding the mgf of $X_1 + X_2$ and the mgf of a $Gamma(2, \theta)$ random variable.