Chapter 1

Statistics: Definitions and Common Distributions

1.1 Definitions

The following definitions are from [**?**]:

Definition 1.1.1 (Population)**.** *A population is the collection of all outcomes, responses, measurements or counts that are of interest.*

Definition 1.1.2 (Sample)**.** *A sample is a subset of a population.*

Definition 1.1.3 (Parameter)**.** *A parameter is a numerical description of a population characteristic.*

Definition 1.1.4 (Statistic)**.** *A statistic is a numerical description of a sample characteristic.*

Definition 1.1.5 (Descriptive Statistics)**.** *Descriptive Statistics is the branch of statistics that involves the organization, summarization and display of data.*

Definition 1.1.6 (Inferential Statistics)**.** *Inferential Statistics is the branch of statistics that involves using a sample to draw conclusions about a population.*

For example, consider the set of undergraduates at Brown University, and say I want to know what percent are Red Sox fans. The percent of undergraduates (our *population*) at Brown who are Red Sox fans is a (population) *parameter*. It would be difficult and time-consuming for me to ask each undergraduate at Brown if they are Sox fans; however, consider the set of students in this class – this is a *sample* from our *population*. The percent of students in our class who are Red Sox fans is a *statistic*. The question is: how good of an estimate is this *statistic* for the population *parameter*? This is a problem in *inferential statistics*.

1.2 Common Distributions

1.3 Moments, Probability Functions, Moment Generating Functions

Definition 1.3.1 (CDF). *For a random variable X, the cumulative distribution function or cdf, often denoted by F_X or P_X, is the probability that X takes on values at most x:*

$$F_X(x) \;=\; \textit{Prob}(X \leq x). \tag{1.1}$$

Definition 1.3.2 (PDF). *For a continuous random variable X, the probability density function or pdf, often denoted by p_X or f_X, satisfies*

$$F_X(x) \;=\; \int_{-\infty}^{x} f_X(x)dx. \tag{1.2}$$

Definition 1.3.3 (PMF). *For a discrete random variable X, the probability mass function or pmf, often denoted by p_X or f_X, satisfies*

$$\textit{Prob}(X = x) \;=\; f_X(x). \tag{1.3}$$

If p is a continuous probability distribution, for $k \in \{0, 1, 2, \dots\}$ we define the k^{th} moment by

$$\mathbb{E}\left[x^k\right] \;=\; \int_{-\infty}^{\infty} x^k p(x)dx. \tag{1.4}$$

An alternate notation is that if f_X is the pdf for a continuous random variable X, then

$$\mathbb{E}\left[X^k\right] \;=\; \int_{-\infty}^{\infty} X^k f_X(x)dx. \tag{1.5}$$

If p is a discrete distribution (taking on at most countably many values x_i), we have

$$\mathbb{E}\left[x^k\right] \;=\; \sum_i x_i^k p(x_i). \tag{1.6}$$

One often writes μ_k' for the k^{th} moment, and μ for the first moment (called the mean). One often studies centered moments

$$\mu_k \;=\; \mathbb{E}\left[(x-\mu)^k\right]. \tag{1.7}$$

The second centered moment, called the variance, is denoted σ^2; its square-root is called the standard deviation and denoted σ.

Definition 1.3.4 (MGF)**.** *The moment generating function or mgf of a random variable X with cdf F_X is given by*

$$M_X(t) \;=\; \mathbb{E}\left[e^{tX}\right] \;=\; \int_{-\infty}^{\infty} e^{tx} f_X(x)dx, \tag{1.8}$$

provided that $M_X(t)$ exists for t in a neighborhood of the origin. For a discrete random variable X with pmf f_X we have

$$M_X(t) \;=\; \mathbb{E}\left[e^{tX}\right] \;=\; \sum_i e^{tx_i} f_X(x_i), \tag{1.9}$$

again assuming the series converges for t in a neighborhood of the origin.

1.3.1 Discrete Distributions

Definition 1.3.5 (Bernoulli Distribution)**.** *The Bernoulli Distribution with parameter $p \in [0,1]$ is specified by*

1. *pmf:*

$$\textit{Prob}\,(X = x) \;=\; \begin{cases} p & \textit{if } x = 1 \\ 1-p & \textit{if } x = 0. \end{cases} \tag{1.10}$$

2. *mean $\mu = p$, variance $\sigma^2 = p(1-p)$.*

3. *mgf: $M_X(t) = (1-p) + pe^t$.*

Definition 1.3.6 (Binomial)**.** *The Binomial Distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$ is specified by*

1. *pmf:*

$$\textit{Prob}\,(X = x) = \binom{n}{x} p^x (1-p)^{n-x},\; x \in \{0, 1, \ldots, n\}. \tag{1.11}$$

2. *mean $\mu = np$, variance $\sigma^2 = np(1-p)$.*

3. *mgf: $M_X(t) = [pe^t + (1-p)]^n$.*

Definition 1.3.7 (Discrete Uniform)**.** *The Uniform Distribution with parameter N is specified by*

1. *pmf:*

$$\textit{Prob}\,(X = x) \;=\; \frac{1}{N},\;\; x \in \{1, 2, \ldots, N\} \tag{1.12}$$

2. *mean* $\mu = \frac{N+1}{2}$, *variance* $\sigma^2 = \frac{(N+1)(N-1)}{12}$.

3. *mgf:* $M_X(t) = \frac{1}{N}\sum_{k=1}^{N} e^{kt}$.

Definition 1.3.8 (Geometric)**.** *The Geometric Distribution with parameter* $p \in [0,1]$ *is specified by*

1. *pmf:*
$$\textit{Prob}\,(X = x) \;=\; p(1-p)^{x-1},\;\; x \in \{1, 2, \dots\}. \tag{1.13}$$

2. *mean* $\mu = \frac{1}{p}$, *variance* $\sigma^2 = \frac{1-p}{p}$.

3. *mgf:* $M_X(t) = \frac{pe^t}{1-(1-p)e^t}$ *for* $t < -\log(1-p)$.

Definition 1.3.9 (Hypergeometric)**.** *The Hypergeometric Distribution with parameters* N, M *and* K *is specified by*

1. *pmf:*
$$\textit{Prob}\,(X = x) \;=\; \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}},\;\; x \in \{0, 1, \dots, K\},\; M-(N-K) \le x \le M,\; N, M, K \ge 0. \tag{1.14}$$

2. *mean* $\mu = \frac{KM}{N}$, *variance* $\sigma^2 = \frac{KM}{N}\frac{(N-M)(N-K)}{N(N-1)}$.

Definition 1.3.10 (Negative Binomial)**.** *The Negative Binomial Distribution with parameters* r *and* $p \in [0,1]$ *is specified by*

1. *pmf:*
$$\textit{Prob}\,(X = x) \;=\; \binom{r+x-1}{x} p^r (1-p)^x,\;\; x \in \{0, 1, \dots\}. \tag{1.15}$$

2. *mean* $\mu = \frac{r(1-p)}{p}$, *variance* $\sigma^2 = \frac{r(1-p)}{p^2}$.

3. *mgf:* $M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$ *for* $t < -\log(1-p)$.

Definition 1.3.11 (Poisson)**.** *The Poisson Distribution with parameter* $\lambda \in [0, \infty)$ *is specified by*

1. *pmf:*
$$\textit{Prob}\,(X = x) \;=\; \frac{e^{-\lambda}\lambda^x}{x!},\;\; x \in \{0, 1, \dots\}. \tag{1.16}$$

2. *mean* $\mu = \lambda$, *variance* $\sigma^2 = \lambda$.

3. *mgf:* $M_X(t) = e^{\lambda(e^t - 1)}$.

1.3.2 Continuous Distributions

Definition 1.3.12 (Beta). *The Beta Distribution with parameters $\alpha, \beta > 0$ is specified by*

1. *pdf:*
$$f(x) \;=\; \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \;\; x \in [0,1], \;\; B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{1.17}$$
2. *mean $\mu = \frac{\alpha}{\alpha+\beta}$, variance $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.*
3. *mgf: $M_X(t) = 1 + \sum_{k=1}^{\infty} \prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r} \frac{t^k}{k!}$.*

Definition 1.3.13 (Cauchy). *The Cauchy Distribution with parameter $\theta \in \mathbb{R}$ and $\sigma > 0$ is specified by*

1. *pdf:*
$$f(x) \;=\; \frac{1}{\pi\sigma} \frac{1}{1 + \left(\frac{x-\theta}{\sigma}\right)^2}. \tag{1.18}$$
2. *the mean and standard deviations do not exist.*
3. *the mgf does not exist.*

Definition 1.3.14 (Chi Squared). *The Chi Squared Distribution with parameter $p \in \{1, 2, \dots\}$ is specified by*

1. *pdf:*
$$f(x) \;=\; \frac{1}{\Gamma(p/2)2^{p/2}} x^{\frac{p}{2}-1} e^{-\frac{x}{2}}, \;\; x \geq 0. \tag{1.19}$$
2. *mean $\mu = p$, variance $\sigma^2 = 2p$.*
3. *mgf: $M_X(t) = \left(\frac{1}{1-2t}\right)^{p/2}$ for $t < \frac{1}{2}$.*

Definition 1.3.15 (Exponential). *The Exponential Distribution with parameter $\beta > 0$ is specified by*

1. *pdf:*
$$f(x) \;=\; \frac{1}{\beta} e^{-x/\beta}, \;\; x \geq 0. \tag{1.20}$$
2. *mean $\mu = \beta$, variance $\sigma^2 = \beta^2$.*
3. *mgf: $M_X(t) = \frac{1}{1-\beta t}$, $t < \frac{1}{\beta}$.*

Definition 1.3.16 (Gamma)**.** *The Gamma Distribution with parameters $\alpha, \beta > 0$ is specified by*

1. *pdf:*
$$f(x) \;=\; \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \;\; x \geq 0. \tag{1.21}$$
2. *mean $\mu = \alpha\beta$, variance $\sigma^2 = \alpha\beta^2$.*
3. *mgf: $M_X(t) = \left(\frac{1}{1-\beta t}\right)^{\alpha}$ for $t < \frac{1}{\beta}$.*

If $\alpha = 1$ one obtains the exponential distribution; if $\alpha = \frac{p}{2}$ and $\beta = 2$, one obtains the chi squared distribution.

Definition 1.3.17 (Normal)**.** *The Normal (or Gaussian) Distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ is specified by*

1. *pdf:*
$$f(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \;\; x \in \mathbb{R}. \tag{1.22}$$
2. *mean $\mu = \mu$, variance $\sigma^2 = \sigma$.*
3. *mgf: $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$.*

Definition 1.3.18 (Uniform)**.** *The Uniform Distribution with parameters $a, b \in \mathbb{R}$ is specified by*

1. *pdf:*
$$f(x) \;=\; \frac{1}{b-a}, \;\; x \in [a, b]. \tag{1.23}$$
2. *mean $\mu = \frac{b+a}{2}$, variance $\sigma^2 = \frac{(b-a)^2}{12}$.*
3. *mgf: $M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$.*

If $a = 0$ and $b = 1$, this is a special case of the beta distribution.

Definition 1.3.19 (Weibull)**.** *The Weibull Distribution with parameters $\beta, \gamma > 0$ is specified by*

1. *pdf:*
$$f(x) \;=\; \frac{\gamma}{\beta} x^{\gamma-1} e^{-x^{\gamma}/\beta}, \;\; x \geq 0. \tag{1.24}$$
2. *mean $\mu = \beta^{1/\gamma}\Gamma(1 + \gamma^{-1})$, variance $\sigma^2 = \beta^{2/\gamma}\left[\Gamma(1 + 2\gamma^{-1}) - \Gamma^2(1 + \gamma^{-1})\right]$.*
3. *the mgf exists only for $\gamma \geq 1$; the moments are $\mathbb{E}\left[X^k\right] = \beta^{n/\gamma}\Gamma(1 + n\gamma^{-1})$.*