3.1. Introduction

"How do all these unusuals strike you, Watson?"
"Their cumulative effect is certainly considerable, and yet each of them is quite possible in itself." [Sherlock Homes and Dr. Watson]

Family of Normal Distribution has a pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

- For each distribution, we will investigate
  - Moments such as Mean and Variance
  - Moment Generating Function
  - Relationship with other distributions

3.2. Discrete Distribution

A. Discrete Uniform

$$X \sim \text{Discrete Uniform}(1, N)$$

$$f_X(x|N) = P(X = x|N) = \frac{1}{N}, \quad x = 1, 2, \cdots, N$$

$E(X) =$
$E(X^2) =$
$Var(X) =$

$$Y \sim \text{Discrete Uniform}(N_0, N_1)$$

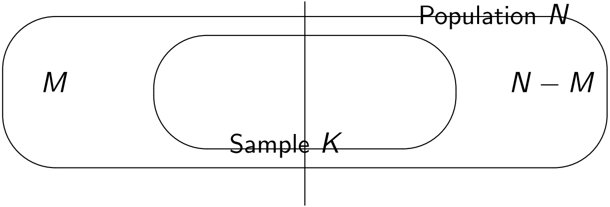
$$f_Y(y|N_0, N_1) = \frac{1}{N_1 - N_0 + 1}, \quad y = N_0, \cdots, N_1$$

$E(Y) =$
$Var(Y) =$

3.2. Discrete Distribution

B. Hypergeometric Distribution

: Related with a single random sample w/o replacement



- Total N objects with 2 groups. One has M elements.
- Select a sample of size K: $\binom{N}{K}$ possible samples.
- Let X be the number of group 1 elements in the sample

3.2. Discrete Distribution

B. Hypergeometric Distribution

$$X \sim \text{Hypergeometric}(N, M, K)$$

$$P(X = x|N, M, K) = f_X(x|N, M, K) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$$

$$x = 0, 1, \cdots, K.$$

$E(X) =$

$Var(X) =$

- Sequence of Bernoulli Trials is a sequence of *identical / uncorrelated* trials with two outcomes *Success(S) and failure(F)*.
- Success (or Failure) on i-th trial, $i = 1, \cdots$, are assumed to be independent.
- $P(S \text{ on trial } i\ ) = p$, $i = 1, 2, \cdots$
- Bernoulli, Binomial, Geometric, Negative Binomial distributions

Consider a single trial. Define X as

$$X = \begin{cases} 1, & \text{if the trial is a success,} \\ 0, & \text{if the trial is a failure.} \end{cases}$$

$$X \sim \text{Bernoulli}(p)$$

$$f_X(x|p) = p^x(1-p)^{1-x}, \quad x = 0, 1.$$

$E(X) =$

$Var(X) =$

3.2. Discrete Distribution

D. Binomial distribution

Let X be the number of success in the first n Bernoulli trials.

$$X \sim \text{Binomial}(n, p)$$

$$f_X(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots, n$$

$E(X) =$

$Var(X) =$

$M_X(t) =$

$\triangleright$ Example: Newly manufactured widgets adopt a Bernoulli process model with $p = 0.01$ for whether these widgets fail a functional test;

• In 500 of these tests, what is the probability all widgets pass ?

$X =$ Number of widgets not functional

$X \sim \text{Binomial}(n = 500, p = 0.01)$

$f_X(\ ?\ | n = 500, p = 0.01) =$

$\triangleright$ Example -Continued

• Now assume p is unknown and 20 defective are observed. What p is the outcome $x = 20$ most likely?

$$f_X(20|n = 500, p = ?) = \binom{500}{20} p^{20}(1-p)^{480}$$

3.2. Discrete Distribution

E. Geometric distribution

Let X be the trial on which the first success occurs or the number of trials until the first success.

$$X \sim \text{Geometry}(p)$$

$$f_X(x|p) = (1-p)^{x-1}p, \quad x = 1, 2, \cdots$$

Note that $\sum_{x=1}^{\infty}(1-p)^{x-1} = 1/p$.
$E(X) =$

$E[X(X-1)] =$

$Var(X) =$

3.2. Discrete Distribution

E. Geometric distribution

- $F_X(x|p) = P[X \leq x|p] = 1 - P[X > x|p]$

- For a given x_0, the conditional probability of the remaining waiting time to the success given that we've waited to x_0 without seeing a success is
$P[X = x_0 + x|X > x_0] =$

- $M_X(t) =$

$\triangleright$ Example: The same example in the binomial distribution with $p = 0.01$. What is the probability of running at least 50 units without a test failure ?

$X=$ number of trials until the first test failure $\sim$ Geometry(p).

$P[X > 50] =$

Y = number of test failure among the first 50 units. Then Y has Binomial distribution thus the probability is

$P[Y = y] = \binom{50}{y} p^y (1-p)^{50-y}$

3.2. Discrete Distribution

F. Negative Binomial distribution

$X =$ trials on which the r-th success occurs.

$$X \sim \text{Negative Binomial}(r, p)$$

$$f_X(x|r,p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \cdots \ .$$

- If $r = 1$ then X has a geometric distribution.
- $Y = X - r =$ number of failure before r-th success

$$f_Y(y|r,p) =$$

$E(X) =$, $Var(X) =$

$M_X(t) =$

Model for the number of occurrences of a relatively rare phenomenon across a fixed interval of time or area of space.

$$X \sim \text{Poisson}(\lambda)$$

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

$E(X) =$, $Var(X) =$

$M_X(t) =$

$\triangleright$ Example: A certain type of tree has seedlings randomly dispersed in a large area, with the mean density of seedlings being approximately five per square yard. If X is the number of such seedlings in 0.25 square yards, then

$$X \sim \text{Poisson}(\lambda = \qquad\qquad)$$

$P[X = 3] =$
$P[X \geq 4] =$
• Recursive:

$$P[X = x] = \qquad\qquad\qquad\qquad = (\lambda/x)P[X = x - 1]$$

• $X_1 \sim \text{Poisson}(\lambda_1)$, $X_2 \sim \text{Poisson}(\lambda_2)$, and X_1 and X_2 are independent. Then $Y = X_1 + X_2$ has

3.3. Continuous Distribution

A. Uniform distribution

$$X \sim \text{Unif}(a, b)$$

$$f_X(x|a, b) = \frac{1}{b-a}, \quad a < x < b.$$

$E(X) =$

$Var(X) =$

$M_X(t) =$

- If $U \sim \text{Unif}(0,1)$, then $X = a + (b-a)U$ has uniform distribution on (a, b).

$$X \sim \text{Gamma}(\alpha, \beta)$$

$$f_X(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \; x > 0, \; \alpha > 0, \; \beta > 0.$$

α: Shape parameter β: Scale parameter

- Γ function

1. $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt, \quad \alpha > 0$
2. $\Gamma(\alpha+1) = \int_0^{\infty} t^{\alpha} e^{-t} dt = \alpha \int_0^{\infty} t^{\alpha-1} e^{-t} dt = \alpha\Gamma(\alpha)$
3. $\Gamma(1) = 1$
4. $\Gamma(n) = (n-1)!$, for positive integer n.

3.3. Continuous Distribution

B. Gamma distribution: [Example 3.3.1 Gamma - Poisson Relationship]

Let $X \sim \text{Gamma}(\alpha, \beta)$, $\quad Y \sim \text{Poisson}(x/\beta)$. Then we have

$$P(X \leq x|\alpha, \beta) = P(Y \geq \alpha|\beta)$$

(See Example 3.1.1 for the recursive calculation.)

$$EX = \alpha\beta \quad , \quad Var(X) = \alpha\beta^2$$

$$M_X(t) = \frac{1}{(1-\beta t)^\alpha}, \quad t < \frac{1}{\beta}$$

- Special gamma distribution

1. $X \sim \chi^2(p)$: χ^2 distribution with df p.

$$\chi^2(p) \equiv \text{Gamma}\left(\alpha = \frac{p}{2}, \beta = 2\right)$$

$EX =$

$Var(X) =$

$M_X(t) =$

2. $X \sim \text{Exp}(\beta)$: Exponential distribution.

$$\text{Exp}(\beta) \equiv \text{Gamma}\,(\alpha = 1, \beta)$$

$$F_X(x|\beta) = 1 - \exp[-x/\beta], \quad x > 0$$

$EX =$

$Var(X) =$

$M_X(t) =$

3.3. Continuous Distribution

B. Gamma distribution: Exponential Distribution

: Used to describe the distribution of time required for the first event

: Memoryless properties

For a give time a,

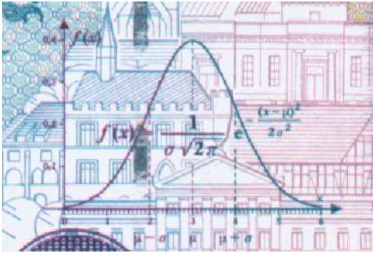
$$P(X > a + t | X > a) = P(X > t).$$

Conditional probability of waiting additional time t after waiting a is the same as the unconditional probability of waiting t.

(See page 101.)

3.3. Continuous Distribution

C. Normal distribution



3.3. Continuous Distribution

C. Normal distribution

One of the most important distribution

$$X \sim N(\mu, \sigma^2)$$

$$f_X(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

$$-\infty < x < \infty, \; -\infty < \mu < \infty, \; \sigma^2 > 0$$

$$1 \overset{?}{\equiv} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

Or, equivalently, (by setting $z = (x-\mu)/\sigma$)

$$1 \overset{?}{\equiv} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right] dz$$

(Use polar coordinate.)

$$M_X(t) = \exp\left[\mu t + \frac{1}{2}\sigma^2 t^2\right], \quad EX = \qquad\qquad , \quad Var(X) =$$

- Standard normal distribution: $N \sim (0, 1)$

If $X \sim (\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] dt \\ &= \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right] dz = \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

• Often used as an approximated distribution of a certain RV. For example, let $X \sim$ Binomial(n, p). *then under the suitable conditions*, X is approximately distributed as $N(\mu = np, \sigma^2 = np(1-p))$. Let $n = 20$ and $p = 0.5$. Then $X \sim$ Binomial$(20, 0.5)$ can be approximated by $Y \sim N(10, 5)$.

$$P(X \leq 12) = \sum_{x=0}^{12} \binom{20}{x} (0.5)^{20} = 0.8684.$$

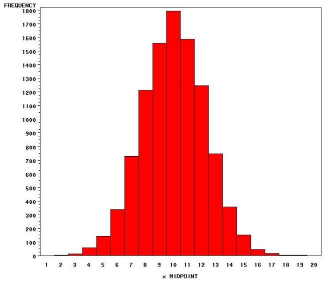
$$P(Y \leq 12) = P\left(Z \leq \frac{12-10}{\sqrt{5}}\right) = 0.8133.$$

Not so good approximation $\rightarrow$ Need continuity correction

$$P(X \leq 12) = P(X \leq 12.5) \approx P(Y \leq 12.5) = 0.8686.$$

3.3. Continuous Distribution

C. Normal distribution



10,000 generated binomial values:
Mean=10.02 and Variance=4.99

- $\chi^2(1)$ can be obtained from $N(0,1)$: Let $Y = Z^2$, where $Z \sim N(0,1)$. Then

$$\begin{aligned} F_Y(y) =& P\left(Z^2 \leq y\right) \\ =& P\left(-\sqrt{y} \leq Z \leq \sqrt{y}\right) \\ =& F_Z(\sqrt{y}) - F_Z(-\sqrt{y}) \end{aligned}$$

Then,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_Z(\sqrt{y})/(2\sqrt{y}) + f_Z(-\sqrt{y})/(2\sqrt{y})$$

$$X \sim \text{Beta}(\alpha, \beta)$$

$$f_X(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1,\ \alpha > 0,\ \beta > 0.$$

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\ .$$

$EX^r =$

$$EX = \qquad\qquad , Var(X) =$$

If $\alpha = \beta = 1$ then $X \sim U(0,1)$

3.4. Exponential Family

Definition

A family pdf pdf (or pmf) is called an exponential family if it can be expressed as

$$f_X(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left[\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right],$$

where $h(x)$, $t_1(x), \cdots, t_k(x)$ are real-valued function of x alone, $c(\boldsymbol{\theta})$, $w_1(\boldsymbol{\theta}), \cdots, w_k(\boldsymbol{\theta})$ are real valued function of $\boldsymbol{\theta}$ only.

3.4. Exponential Family

- ▷ Example: $X \sim$ Binomial(n, p). Known n.

$$f_X(x|p) = \binom{n}{x} p^x (1-p)^{(n-x)}$$

- ▷ Example: $X \sim N(\mu, \sigma^2)$.

$$f_X(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

3.4. Exponential Family

- ▷ Example: $X \sim \text{Geometric}(p)$.

$$f_X(x|p) = p(1-p)^{x-1}$$

- ▷ Example: $X \sim \text{Gamma}(\alpha, \beta)$.

$$f_X(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

3.5. Location & Scale family

Theorem

Let $f(x)$ be any pdf and let $\mu \in \mathcal{R}$ and $\sigma > 0$ be given constants. Then the function

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

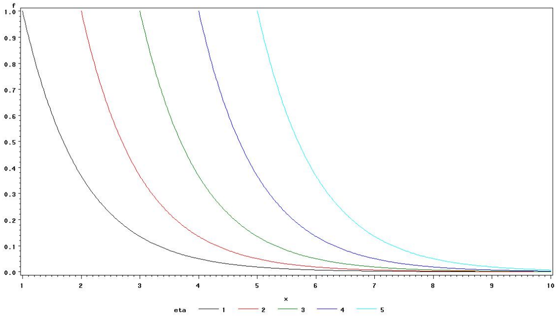
is a pdf.

Definition

Let $f(x)$ be any pdf. Then the family of pdfs $f(x-\mu)$, $-\infty < \mu < \infty$, is called the *location family* with standard pdf $f(x)$ and μ is called the *location parameter* for the family.

3.5. Location & Scale family

▷ Example: $f(x|\eta) = \exp[-(x-\eta)],\ x > \eta$



Used in life-testing application.
η is often called threshold parameter.

3.5. Location & Scale family

• Location parameter is often related with a measure of central tendency of distribution.

$\triangleright$ Example: $X \sim N(\mu, 1)$

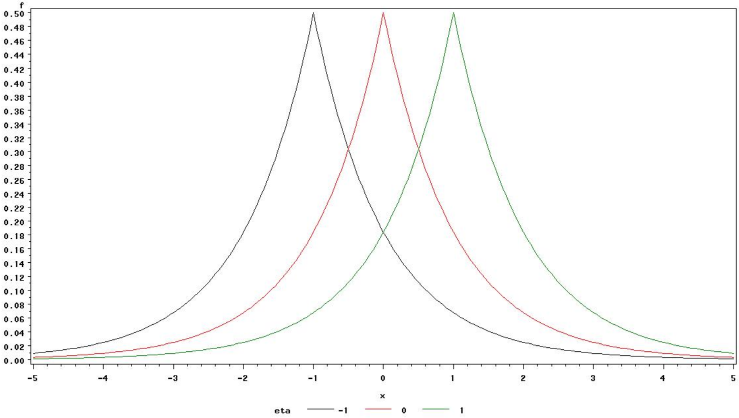
$$f_X(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x-\mu)^2\right]$$

$\triangleright$ Example:

$$f_X(x|\mu,\sigma^2) = \frac{1}{2} \exp\left[-|x-\eta|\right]$$

$\mu,\ \eta$ are mean, mode, median of distribution

3.5. Location & Scale family



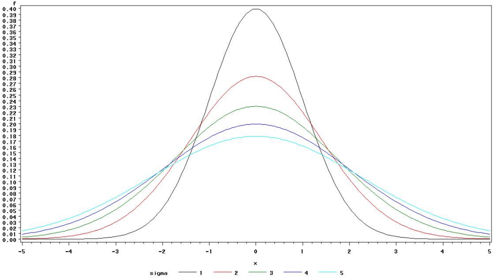
3.5. Location & Scale family

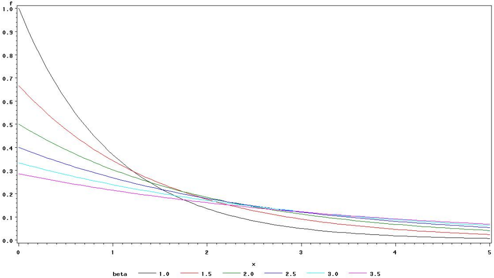
Definition

Let $f(x)$ be any pdf. Then the family of pdfs $(1/\sigma)f(x/\sigma)$, $0 < \sigma$, is called the *scale family* with standard pdf $f(x)$ and σ is called the *scale parameter* for the family.

• The scale parameter σ either stretches $(\sigma > 1)$ or contracts $(\sigma < 1)$ the graph $f(x)$ while still maintaining the basic shape of the distribution.
$\triangleright$ Examples: $X \sim N(0, \sigma^2)$, $X \sim \exp(\beta)$

3.5. Location & Scale family





3.5. Location & Scale family

Definition

Let $f(x)$ be any pdf. Then the family of pdfs $(1/\sigma)f[(x-\mu)/\sigma]$, $-\infty < \mu < \infty$, $0 < \sigma$, is called the *location-scale family* with standard pdf $f(x)$ and μ is called the *location parameter* and σ is called the *scale parameter* for the family.

$\triangleright$ Example: Standard distribution: $f(x) = \exp(-x)$, $x > 0$.

$$\text{Location family: } \; e^{-(x-\eta)}, \quad \eta > 0, \quad x > \eta$$

$$\text{Scale family: } \; \frac{1}{\beta} e^{-x/\beta}, \quad \beta > 0, \quad x > 0$$

$$\text{Location-Scale family: } \; \frac{1}{\beta} e^{-(x-\eta)/\beta}, \quad \eta > 0, \quad \beta > 0, \quad x > \eta$$

3.5. Location & Scale family

Theorem

Let $f(\cdot)$ be any pdf. Let μ be any real number and let σ be any positive real number. Then X is a random variable with pdf

$$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

if and only if there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$.

3.5. Location & Scale family

Theorem

Let $Z \sim f(z)$ and assume $E(Z)$ and $Var(Z)$ exist. If X is a RV with pdf

$$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

then

$$E(X) = \sigma E(Z) + \mu, \quad Var(X) = \sigma^2 Var(Z).$$

If $E(Z) = 0$ and $Var(Z) = 1$ then $EX = \mu$ and $Var(X) = \sigma^2$.

Theorem (Chebychev's Inequality)

Let X be a random variable and let $g(x)$ be a nonnegative *function. Then, for any $r > 0$,*

$$P\left[g(X) \geq r\right] \leq \frac{E\left[g(X)\right]}{r} \quad .$$

3.6. Inequalities and Identities

3.6.1. Probability Inequality

Theorem

If $P[Y \geq 0] = 1$ *and* $P[Y = 0] < 1$, *then for any* $r > 0$

$$P[Y \geq r] \leq \frac{EY}{r}$$

with equality if and only if

$$P[Y = r] = p = 1 - P[Y = 0], \quad 0 < p \leq 1.$$

3.6. Inequalities and Identities

3.6.2. Identity [Lemma 3.6.5 Stein's Lemma]

Lemma

$X \sim N(\mu, \sigma^2)$. Let g be a differentiable function satisfying $E[|g'(x)|] < \infty$. Then

$$E[g(X)(X-\mu)] = \sigma^2 E[g'(X)].$$

$EX^3 =$