Distributions

Common Distributions

a) Uniform

(i)
$$f(x) = \frac{1}{b-a}, a < x \le b$$
$$F(x) = \frac{(x-a)}{b-a}, a < x \le b \text{ (discrete or continuous)}$$

ii) Matlab: rand(1) returns random number between 0 and 1 with uniform distribution

(1) Useful for simulations p = 0.5 draw number, event occurs if number < 0.5.

b) Bernoulli

i) Single event, two outcomes

$$\mathrm{P}(sucess) = p; \mathrm{P}(failure) = 1 - p$$

(i)
$$\mathrm{P}(k) = \begin{Bmatrix} p & \text{k = 1} \\ 1-p & \text{k = 0} \end{Bmatrix}$$

(ii) Single trial behavior example

Animal put in middle of maze, has some bias

$$F(choice) = f(choice) = \begin{Bmatrix} 0.1 & right \\ 0.9 & left \end{Bmatrix}$$

c) Binomial

i) Line up n independent Bernoullis, probability of k successes.

$$\mathrm{P}(sucess) = p; \mathrm{P}(failure) = 1 - p$$

(1) Model for behavior for a well trained animal or number of spikes from a constant rate cell

ii) Enumeration for n = 3

$$f(k) = \mathrm{P}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

(i)
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$
$$F(k) = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

Matlab function for combinations: NCHOOSEK

Example:

Draw approximate f(x) for 10 trials, p(1) = .5 and p(1) = .1

d) Gaussian distribution

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

i) where σ is the standard deviation

μ is the mean.

$F(x)$ is roughly sigmoidal

Example:

e) Exponential

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where λ is the rate parameter

Notation: $X \sim \mathrm{Exp}(\lambda)$.

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Steps in testing a hypothesis about data

a) Characterize the data
   i) Calculate measures of mean, variance, etc.
   ii) Estimate distribution
b) Compare parameters or full distribution to reference distribution (Computed or calculated)
c) Example problem: given a set of samples $a_1, a_2, ..., a_n$ from distribution A and another set $b_1, b_2, ..., b_n$ from distribution B, how do we determine whether A and B differ?
   i) Probability of each particular observation is low for discrete process and 0 for continuous process.
   ii) We need to estimate the parameters of distributions A and B from the data and then compare those parameters and their error estimates.

Measures

a) Mean (Expected Value)

(i)
$$\mathrm{E}[X] = \sum_{i=-\infty}^{\infty} xf(x) \quad \text{for discrete distributions}$$
$$\mathrm{E}[X] = \int_{-\infty}^{\infty} xf(x)\,dx \quad \text{for continuous distributions}$$

(ii) Corresponds to center of mass

ii) Properties of expected value

(1) Linearity $\mathrm{E}[aX+b] = a\mathrm{E}[X]+b$ (show)

(2) $\mathrm{E}[X_1+...+X_n] = \mathrm{E}[X_1]+...+\mathrm{E}[X_n]$

(3) If $X_1,...,X_n$ are independent random variables, $E\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n}\mathrm{E}[X_i]$

b) $\mathrm{Var}(X) = \sigma^2 = E\left[(X-\mu)^2\right]$ where $\mu = \mathrm{E}[X]$

i) Poperties of variance

(1) $\mathrm{Var}[aX+b] = a^2\mathrm{Var}[X]$

(2) $\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - \mathrm{E}\left[X\right]^2$

(3) If $X_1,...,X_n$ are independent random variables,

$\mathrm{Var}[X_1+...+X_n] = \mathrm{Var}[X_1]+...+\mathrm{Var}[X_n]$

ii) σ is the standard deviation

c) $\mathrm{Median}(X) = m =$ point where $\Pr(X \le m) \ge \frac{1}{2}$ and $\Pr(X \ge m) \ge \frac{1}{2}$.

d) $\mathrm{Mode}(X) =$ argmax $f(x)$

Returning to the distributions we defined above

a) Uniform

$f(x) = \frac{1}{b-a}, a \le x \le b$

$\mathrm{E}[x] = \frac{a+b}{2}$

$\mathrm{Var}[X] = \frac{(b-a)(b-a+2)}{12}$ (discrete)

$\mathrm{Var}[X] = \frac{(b-a)^2}{12}$ (continuous)

b) Bernoulli

$$\Pr(k) = \begin{cases} p & \mathrm{k} = 1 \\ 1-p & \mathrm{k} = 0 \end{cases}$$

$$\mathrm{E}[\mathrm{k}] = p$$

$$\mathrm{Var}[k] = p(1-p)$$

c) Binomial

$$\Pr(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mathrm{E}[X] = np$$

$$\mathrm{Var}[X] = np(1-p)$$

Question: how would you derive mean and variance from Bernoulli?

d) Gaussian

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

$$\mathrm{E}[x] = \mu$$

$$\mathrm{Var}[x] = \sigma^2$$

e) Exponential

$$E[x] = \frac{1}{\lambda}$$

$$\mathrm{Var}[x] = \frac{1}{\lambda^2}$$

Computing Parameters

a) We often don't have access to the actual distributions and we often don't have enough information to estimate the entire distribution, so we calculate summary statistics.

b) Sample Mean: mean of random samples X_i is $\overline{X} = \frac{1}{n}\left(X_1 + ... + X_n\right)$

(1) Properties of the sample mean:

$$\mathrm{E}\left[\overline{X}\right] = \mu$$

$$\lim_{n\to\infty} \mathrm{E}\left[\overline{X_n}\right] = \mu$$

$$\mathrm{Var}\left(\overline{X}\right) = \frac{\mathrm{Var}(X)}{\mathrm{n}} = \frac{\sigma^2}{\mathrm{n}},$$

$$\sqrt{\mathrm{Var}\left(\overline{X}\right)} = \sqrt{\frac{\sigma^2}{\mathrm{n}}} = \frac{\sigma}{\sqrt{\mathrm{n}}}$$

but....

$\frac{\sigma}{\sqrt{\text{n-1}}}$ is less biased and is used for the standard error of the mean

Note that this does not apply to all distributions.

c) Sample standard deviation: standard deviation of random samples

$$s = \sqrt{\sum_{i=1}^{N} \frac{\left(X_i - \overline{X}\right)^2}{n-1}}$$

i) Note the denominator.

d) Standard Normal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

Z-scores

(1) Number of stdevs from mean

$$z(x) = \frac{x - \overline{X}}{s}$$

Example: