Probability and Statistics Review

Chapter 1

Guani Wu

STATS 102C: Introduction to Monte Carlo Methods

Distribution and Density Functions

- The cumulative distribution function (cdf) of a random variable X is F_X defined by

$$F_X(x) = P(X \leq x),\ x \in \mathbb{R}$$

- In fact, the distribution of x is completely determined by the cdf, regardless of x being discrete or continuous (or mixed).

Expectation

The concept of the expected value of a random variable parallels the notion of a weighted average. That is, the possible values of the random variable are weighted by their probabilities.

- **Definition:** If X is a discrete random variable with frequency function $p(x)$, then

$$E(X) = \sum_i x_i p(x_i)$$

provided that $\sum_i |x_i| p(x_i) < \infty$. If the sum diverges, the expectation is undefined.

- **Definition:** If X is a continuous random variable with density $f(x)$, then

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

provided that $\int |x| f(x) dx < \infty$. If the integral diverges, the expectation is undefined.

Moments and Variance

- Let $\mu_X = E(X)$. Then μ_X is called the first moment of X. The r^{th} moment of X is $E(X^r)$
- **Definition:** If X is a random variable with expected value $E(X)$, the variance of X is

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

  provided that the expectation exists. The standard deviation of X is the $\sqrt{Var(X)}$
- The standard deviation of a random variable is an indication of how dispersed the probability distribution is about its expectation.

Bernoulli distribution

A Bernoulli distribution takes on only two values: 0 and 1, with probabilities $1-p$ and p, respectively.

- pmf: $p(x) = \begin{cases} p^x(1-p)^{1-x}, & \textit{if} \ \ x=0 \ \ \textit{or} \ \ x=1 \\ 0, & \textit{otherwise} \end{cases}$
- cdf: $\begin{cases} 0, & \textit{if} \ \ x<0 \\ 1-p, & \textit{if} \ \ 0 \le x < 1 \\ 1, & \textit{if} \ \ x \ge 1 \end{cases}$
- mean: p
- variance: $p(1-p)$
- parameter: $p \in [0,1]$
- example: toss a coin once, p=probability that head occurs

Binomial distribution

Suppose that n independent Bernoulli trials are performed, where n is a fixed number. The total number of 1 appearing in the n trials follows a binomial distribution with parameters n and p.

- pmf: $p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & if \;\; x = 0, 1, \ldots, n \\ 0, & otherwise \end{cases}$
- cdf: $\sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{(n-i)}$
- mean: np
- variance: $np(1-p)$
- parameter: $p \in [0, 1]$, $n = 1, 2, \ldots$
- example: the number of heads, toss a coin n times

Poisson distribution

Limit of binomial distributions $X_n \sim B(n, p_n)$, where $p_n \to 0$ as $n \to \infty$ in such a way that $\lambda_n \equiv np_n \to \lambda$.

- pmf: $p(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \ldots \\ 0, & otherwise \end{cases}$
- cdf: $e^{-\lambda} \sum_{i=0}^{x} \frac{\lambda^i}{i!}$
- mean: λ
- variance: λ
- parameter: $\lambda > 0$
- example: number of phone calls coming into an exchange during a unit of time

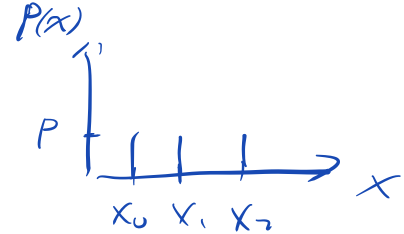
Geometric distribution

The geometric distribution is constructed from an infinite sequence of independent Bernoulli trials. Let X be the total number of trials up to and excluding the first appearance of 1, then X follows the geometric distribution.

- pmf: $p(x) = \begin{cases} (1-p)^x p, & x = 0, 1, 2, \ldots \\ 0, & \textit{otherwise} \end{cases}$
- cdf: $1 - (1-p)^{x+1}$ x=0,1,2,...
- mean: $\frac{1-p}{p}$
- variance: $\frac{1-p}{p^2}$
- parameter: $p \in [0, 1]$
- example: lottery, the number of tickets a person must purchase up to and including the first winning ticket
- a memoryless distribution

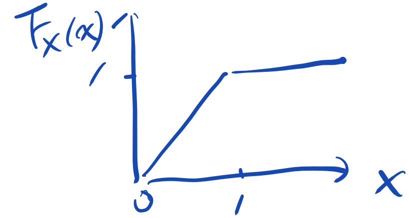
Uniform Distribution

The distribution describes an experiment where there is an arbitrary outcome that lies between certain bounds. The bounds are defined by the parameters, a and b, which are the minimum and maximum values.



- pdf: $f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x < a \;\; or \;\; x > b \end{cases}$

Unif(0, 1)



- cdf: $\begin{cases} 0, & for \;\; x < a \\ \frac{x-a}{b-a} & for \;\; a \leq x \leq b \\ 1, & for \;\; > b \end{cases}$

$$\frac{I(a \leq x \leq b)}{b-a}$$

- mean: $\frac{1}{2}(a+b)$
- variance: $\frac{1}{12}(b-a)^2$
- parameter: $-\infty \leq a \leq b < \infty$

I(statement)

True → 1

False → 0

Exponential Distribution

The exponential distribution is the probability distribution of the time between events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant average rate.

- pdf: $f(x) = \underline{\lambda e^{-\lambda x}}$ for $0 \leq x < \infty$
- cdf:$1 - e^{-\lambda x}$

  λ : call/hour   minutes/call
- mean: $\frac{1}{\lambda}$
- variance: $\frac{1}{\lambda^2}$
- parameter: $\lambda > 0$
- a memoryless distribution
- example: the amount of time (beginning now) until an earthquake occurs, the length (in minutes) of long distance business telephone calls, etc.

let λ be the expected # of calls during a 1-minute interval

If $\lambda = 2$, in one minute, 2 calls

~ two minuts, 4 calls

⋮

in x minutes, $x\lambda$ calls

the number of calls in x minute-interval N_x

$$N_x \sim \text{Poisson}(x \cdot \lambda)$$

Let X be the wait time until the 1st call from any start point

$P(\text{wait at least } x \text{ mintes for the 1st call}) = P(X > x)$

$= P(\text{there was no calls in the 1st } x \text{ minutes})$

$$= P(N_x = 0) = \frac{e^{-\lambda x}(\lambda x)^0}{0!} = e^{-\lambda x}$$

the prob that an event does occur during x minutes

units of time

Beta Distribution

The beta distribution is a family of continuous probability distributions defined on the interval $[0, 1]$ parameterized by two positive shape parameters, denoted by α and β, that appear as exponents of the random variable and control the shape of the distribution.

- pdf: $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ where $0 \le x \le 1$
- cdf: $\frac{\int_0^x t^{\alpha-1}(1-t)^{\beta 1}}{B(\alpha,\beta)}$
- mean: $\frac{\alpha}{\alpha+\beta}$
- variance: $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- parameter: $\alpha > 0$, $\beta > 0$
- example: the beta distribution is the conjugate prior for the Bernoulli, binomial, negative binomial and geometric distributions in Bayesian inference.

$\Gamma(r) = (r-1)!$

$\Gamma(1) = 1$

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$

E.g. $\alpha = 1$, $\beta = 1$

$f(x) = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} x^0 (1-x)^0 = \frac{1}{1} = 1 = \text{Unif}(0,1)$

Estimation Theory

- The (Weak) Law of Large Numbers: Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Let $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$. Then, for any $\varepsilon > 0$.

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$$

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| > \varepsilon) = 0$$

almost surly

- The Mean Squared Error (MSE):

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + [\underbrace{E(\hat{\theta}) - \theta}_{\text{bias}}]^2 \qquad MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

$\hat{\theta}$: estimator

θ : parameter

If $MSE(\hat{\theta})$ converges to 0, then $\hat{\theta}$ is a consistent estimator.

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

$E(\bar{X}) = \mu$ bias$^2 = 0$

$\bar{X} \to \mu$

$Var(\bar{X}) = \frac{\sigma^2}{n}$ $\lim_{n\to\infty} Var(\bar{X}) = 0$

$S^2 \to \sigma^2$

$\bar{X}$ is a consistent estimor of μ

$\hat{F}_X(x) \to F_X(x)$

Empirical Distribution Function (EDF)

- An estimate of $F(x) = P(X \leq x)$ is the proportion of sample points that fall in the interval $(-\infty, x]$. This estimate is called the empirical cumulative distribution function (ecdf) or empirical distribution function.
- Given a value x_0, $F(x_0) = p(X_i \leq x_0)$, for every $i = 1, \ldots, N$.
- $F(x_0)$ is the probability of the event $\{X_i \leq x_0\}$
- $\hat{F}(x_0) = \frac{\sum_{i=1}^{n} \mathbb{I}(x_i \leq x_0)}{n}$
- Is $\hat{F}(x)$ a consistent estimator of $F(x)$?

$\{X_i \leq x_0\}$ $X_1, \ldots, X_i, \ldots, X_n$

$$\hat{F}_n(x_0) = \frac{\#\text{ of } X_i \le x_0}{\text{total \# of obs.}} = \frac{\sum I(X_i \le x_0)}{n} \quad \text{-- ECDF}$$

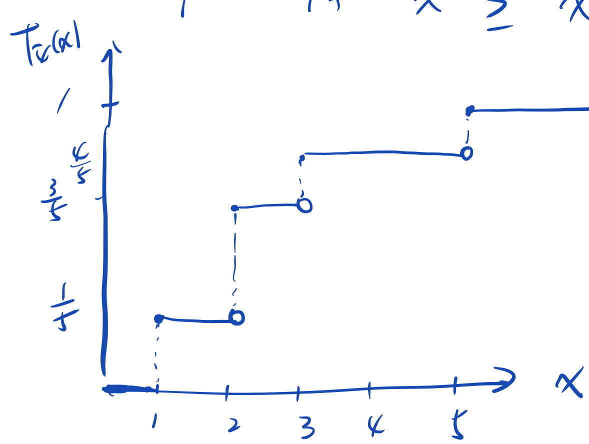
EDF

E.g. Discrete Case: $\underline{1, 2, 2, 3, 5}$

rank: (1) (2) (3) (4) (5)

$$\hat{F}_n(x) = \begin{cases} 0, & x < 1 & x < x_{(1)} \\ \frac{1}{5}, & 1 \le x < 2 & x_{(1)} \le x < x_{(3)} \\ \frac{3}{5}, & 2 \le x < 3 & x_{(3)} \le x < x_{(4)} \\ \frac{4}{5}, & 3 \le x < 5 & x_{(4)} \le x < x_{(5)} \\ 1, & x \ge 5 & x \ge x_{(5)} \end{cases}$$

$$\hat{F}_n(x) = \begin{cases} 0 & \text{if } x < x_{(1)} \\ \frac{i}{n} & \text{if } x_{(i)} \le x < x_{(i+1)} \\ 1 & \text{if } x \ge x_{(5)} \end{cases}$$



population CDF

$$F_X(x_0) = \frac{\sum_{i=1}^{N} I(X_i \le x_0)}{N}$$

set $Y = \underline{I(X_i \le x_0)}$ $\quad Y_i = \begin{cases} 1 & \text{if } X_i \le x_0 \\ 0 & \text{if } X_i > x_0 \end{cases}$

So for given x_0, $\underline{Y_i} \sim Ber(F_X(x_0)) \leftarrow$

$E(Y_i) = F_X(x_0)$

$Var(Y_i) = F_X(x_0)\,[1 - F_X(x_0)]$

$$MSE(\underline{\hat{F}_X(x_0)}) = \underline{Var[\hat{F}_X(x_0)]} + [E(\underline{\hat{F}_X(x_0) - F_X(x_0)})]^2$$

$$\hat{F}_n(x_0) = \frac{\sum_{i=1}^{n} I(X_i \le x_0)}{n} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

$$E(\hat{F}_n(x_0)) = E\left[\frac{\sum Y_i}{n}\right] = E(Y_i) = F_X(x_0) \rightarrow \text{unbiased}$$

$$Var[\underline{\hat{F}_X(x_0)}] = \underline{\underline{\frac{F_X(x_0)\cdot[1 - F_X(x_0)]}{n}}} \le \frac{1}{4n}$$

$$\lim_{n \to 0} Var[\hat{F}_X(x_0)] = 0$$

By CLT

$$\sqrt{n}\,[\underline{\hat{F}_n(x_0) - }\,F(x_0)] \sim N\{0, F_X(x_0)\cdot[1 - F_X(x_0)]\}$$