Bayesian Inference
Useful Distributions

UBCO MDS — DATA 582

Notation

- Denote random variables (RV) by capitals X and observed instances by their lower case x.
- For discrete RVs we have probability mass functions (or *pmf*s), $p(x)$, that give the probability that X is equal to some value, i.e.

$$P(X = x) = p(x)$$

- For continuous RVs we have probability density functions (or *pdf*s). Rather than $p(x)$ directly giving us the probabilities, we must integrate $p(x)$ to give us the probability that X falls within some *interval*:

$$P(a \leq X \leq b) = (a < X < b) = \int_a^b p(x)dx$$

Notation

- For continuous random variables, we can view probability as area under the curve.
- For an infinitesimal (very small) range dx, we can calculate this probability geometrically using the $\textit{base} \times \textit{height}$:

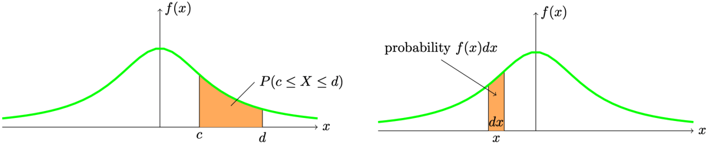

Figure: Image source

Recall that $P(X = x) = 0$ for all values of x when X is a continuous random variable.

Discrete Random Variables

Discrete Uniform Distribution

A random variable X has a *discrete uniform distribution* if and only if each possible value of X are equally likely, that is, it has the probability distribution as

$$f(x) = \frac{1}{k}, \qquad \text{for } x = x_1, x_2, \ldots, x_k$$

where $x_i \neq x_j$ when $i \neq j$.

Eg. Rolling a die.

Bernoulli Distribution

- Suppose an experiment has only two possible outcomes: 1 ("success") and 0 ("fail"), with probabilities of p and $1 - p$ respectively.
- We define a random variable X that takes values of 1 and 0 with probability p and $1 - p$.
- Then, we say X follows a *Bernoulli distribution.* with parameter p.

Eg. Flipping a coin and checking if it's heads.

Bernoulli Distribution

A random variable X has a *Bernoulli distribution*, denoted by $X \sim \text{Bernoulli}(p)$, if and only if it has probability distribution function as

$$f(x) = p^x(1-p)^{1-x} \qquad \text{for } x = 0, 1$$

where $p = P(X = 1)$.

Binomial Distribution

- A so-called Bernoulli trial concerns the outcome of a <u>single</u> experiment having two possible outcomes.
- If we consider n independent repetitions of this trial and define X to count the number of success, X is said to follow a *Binomial* distribution with parameters (n, p).

Eg. Flipping a coin n times and counting the number of heads.

Binomial Distribution

A random variable X has a *binomial distribution*, denoted by $X \sim \mathrm{Bin}(n, p)$ if and only if it has the probability distribution function

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

for $x = 0, 1, 2, \ldots, n$.

Negative Binomial Distribution

- In connection with the repeated Bernoulli trial, we may be interested in the total number of trials in order to have k successes.
- If X counts the number of trials up to and including the k success it is said to follow a *Negative Binomial* distribution with parameters (k, p).

Eg. Counting the number of times you flip a coin until you see k heads.

Negative Binomial

A random variable X has a *negative binomial distribution*, denoted by $X \sim \mathrm{NB}(k, p)$ if and only if it has the probability distribution function

$$f(x; k, p) = \left(\begin{array}{c} x-1 \\ k-1 \end{array} \right) p^k (1-p)^{x-k}$$

for $x = k, k+1, k+2, \ldots$.

Geometric Distribution

- In connection with the repeated Negative Binomial the geometric distribution counts the number of Bernoulli trials until the first success.
- In other words, the special case of the NB where $k = 1$ is said to follow a *Geometric* distribution with parameter p.

> Eg. Counting the number of times you flip a coin until you see the first head.

Geometric Distribution

A random variable X is said to have a *geometric probability distribution*, denoted by $X \sim \mathcal{G}(p)$, if and only if it has the probability distribution function

$$f(x) = p(1-p)^{x-1}$$

for $x = 0, 1, 2, 3, \ldots$
where $0 \leq p \leq 1$.

Poisson Distribution

- The Poisson distribution is used to model counts.
- Usually, the observation process is considered to be taking place over a fixed interval of time (or space) when events occur randomly and independently with constant rate λ.

Eg. number of fish caught in a hour of ice fishing.
Eg. number of flaws on a piece of fabric.

Poisson Distribution

A random variable X has a Poisson distribution, denoted by $X \sim$Poisson(λ), if and only if it has the pmf

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

for $x = 0, 1, 2, 3, \ldots$.
$\lambda > 0$

Continuous Random Variables

Continuous Uniform Distribution

- Similar to the idea of discrete uniform distribution, given the range of a continuous random variable X from α to β for $\alpha < \beta$.
- If all the possible values of X are equally likely in the sense of equal probability density, then we have the plot of probability density vs X as shown on the following slide.
- For continuous RV, probabilities correspond to areas under the curve.

Continuous Uniform Distribution

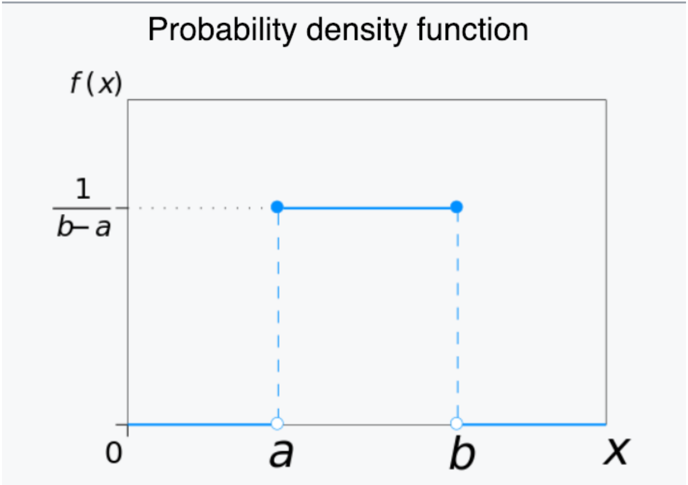

Photo from Wikipedia

Continuous Uniform Distribution

A random variable has a uniform distribution, denoted by, $X \sim \text{unif}(\alpha, \beta)$, if and only if it has the pdf

$$f(x; \alpha, \beta) = \frac{1}{\beta - \alpha},$$

for $\alpha < x < \beta$.

Eg. Assuming a clock represents the real numbers 0–12, spin the minute hand and seeing where points to (an example of an outcome: 2.098712346).

Gamma Distribution

- For a special case of the Gamma distribution (when α is an integer k) is the continuous analogue of the negative binomial distribution.

> Eg. time until you catch your 5th fish.
> Eg. area of fabric until you see the 2nd flaw.

Gamma Distribution

- If X is the length of time from the beginning of the experiment until the kth event X is said to follow a Gamma distribution with parameters α and β.
- $\alpha = k$ and $\beta = 1/\lambda$, where λ is the rate from the Poisson (α need not be an integer, but it has a nice interpretation when it is.)

Gamma Distribution

A random variable X has a gamma distribution , denoted by $X \sim \text{Gamma}(\alpha, \beta)$, if and only if it has the pdf

$$f(x; \alpha, \beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \qquad \text{for } x > 0,$$

where $\alpha > 0$ and $\beta > 0$. The gamma function is defined as:

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy, \qquad \text{for } y > 0.$$

Gamma Distribution: Exponential case

The gamma distribution is a *family* of distributions that includes the exponential and chi-square distributions.

- The RV $X \sim \mathit{Exp}(\beta = 1/\lambda)$ is the length of time from the beginning of the experiment until the *first* event. (It is the continuous analogue of the geometric distribution.)
- In addition, $X \sim \mathit{Exp}(\beta)$ describes the **waiting time** between events in a Poisson process (i.e. the length of time between two successive events has the same exponential distribution.)

Exponential Distribution

A random variable X has *exponential distribution*, denoted by $X \sim \text{Exp}(\beta)$, if and only if its pdf is in the form

$$f(x; \beta) = \frac{1}{\beta} e^{-x/\beta}, \qquad \text{for } x > 0$$

where $\beta > 0$,

Exponential Distribution

Alternative parameterization: $X \sim \text{Exp}(\lambda)$, if and only if its pdf is in the form

$$f(x; \lambda) = \lambda e^{-\lambda x}, \qquad \text{for } x > 0$$

where $\lambda = 1/\beta > 0$. Straightforwardly, the mean and the variance of the exponential distribution are

$$\mu = \beta = \frac{1}{\lambda}, \qquad \text{and} \qquad \sigma^2 = \beta^2 = \frac{1}{\lambda^2}$$

Gamma Distribution: Chi-squared case

- The chi-square distribution is another special case of gamma distribution with $\alpha = \nu/2$, $\beta = 2$.
- The parameter ν is referred to as the number of degrees of freedom (df).

Eg. the sum of the squares of *k* independent standard normal random variables.

Chi-squared Distribution

A random variable X has a chi-square distribution with ν df, denoted by $X \sim \chi^2_\nu$, if and only if it has the pdf as

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu}{2}-1} e^{-x/2},$$

for $x > 0$.

Beta Distribution

- A binomial distribution models the number of successes, X, among n trials given a constant probability of success p for each trial.
- If a probability p is not a constant but a random variable that its value could be different, what would be an appropriate distribution to model the distribution of p?
- Answer: the beta distribution is usually a good choice.

Beta Distribution

A random variable X has a beta distribution if and only if it has the pdf

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

for $0 < x < 1$, where $\alpha > 0$, and $\beta > 0$ and the *beta function* given by

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

Normal Distribution

- The normal distribution—also referred to as the *Gaussian Distribution*—is the most widely used continuous probability distribution in statistics.
- A normal distribution has a bell-shaped probability density function with two parameters: mean μ and the variance σ^2.
- Its mean, μ, is located right at the centre of its density curve.

Normal Distribution

A random variable X is said to have a *normal distribution*, denoted by $X \sim N(\mu, \sigma^2)$ if and only if it has pdf in the form of

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

for $-\infty < x < \infty$, where $\sigma > 0$.