probability space

It's a triplet $(\Omega, \mathcal{F}, P)$

- Ω: A nonempty set, each element ω of Ω is called an *outcome* and Ω is called the *sample space*. The number of ω is called the cardinality of Ω
- $\mathcal{F}$: read as *Script F*, a set of all subsets of Ω, also call it *events*.
- P: a *probability measure on F*. $P(A)$ is the probability of event A ($A \in \mathcal{F}$).

We use $\complement_A$ or A^C to mean the complement of A.

Event axioms:

- Ω is an event ($\Omega \in \mathcal{F}$).
- If A is an event then A^C is an event ($A \in \mathcal{F} \implies A^C \in \mathcal{F}$).
- If A and B are events then $A \cup B$ is an event ($A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$).

Probability axioms:

- $\forall A \in \mathcal{F}, P(A) \geq 0$.
- if $A, B \in \mathcal{F}$ and A and B are mutually exclusive , then $P(A \cup B) = P(A) + P(B)$.
- $P(\Omega) = 1$

Calculate the size of various sets

Principle of counting: If there are m ways to select one variable and n ways to select another varaible, and if these two selections can be made independently, then there is a total of mn ways to make the pair of selections.

n choose k: $\binom{n}{k}$

A random variable is a real-valued function on Ω:

- *pmf* (probability mass function):$p_X(u) = P\{X = u\}$
- $P\{X \in \{u_1, u_2 ...\}\} = \sum_i p_X(u_i) = 1$

The mean of a random variable:
The *mean* (also called *expectation*) of a random variable X with *pmf* p_X is denoted by $E[X]$ and is defined by $E[X] = \sum_i u_i p_X(u_i)$, where $u_1, u_2, ...$ is the list of possible values of X.

The general formula for the mean of a function, $g(X)$, of X, is $E[g(X)] = \sum_i g(u_i) p_X(u_i)$.
$E[ag(X) + bh(X) + c] = aE[g(X)] + bE[h(X)] + c$.

The variance and standard deviation of a random variable:
The *variance* of a random variable X is a measure of how spread out the pmf of X is. Letting $\mu_X = E[X]$, the variance is defined by: $Var(X) = E[(X - \mu_X)^2] = E[(X - E[X])^2] = E[X^2] - 2\mu_X E[X] + \mu_X^2 = E[X^2] - \mu_X^2$
$E[aX + b] = aE[X] + b$, $Var(aX + b) = a^2 Var(X)$
The *standardized version* of X is the random variable $\dfrac{X - \mu_X}{\sigma_X}$, and $Var\left(\dfrac{X - \mu_X}{\sigma_X}\right) = 1$

Conditional probabilities

The **conditional probability** of B given A is defined by: $P(B|A) = \begin{cases} \dfrac{P(AB)}{P(A)} & \text{if } P(A) > 0 \\ \text{undefined if } P(A) = 0 \end{cases}$

Mutually independent events

Event A is **independent** of event B if $P(AB) = P(A)P(B)$.

Events A,B and C are **pairwise independent** if $P(AB) = P(A)P(B), P(AC) = P(A)P(C), P(BC) = P(B)P(C)$

Events A,B and C are **independent** if ther are *pairwise independent* and if $P(ABC) = P(A)P(B)P(C)$

Discrete-type indepent random variables

Random variables X and Y are **independent** if any event of the form $X \in A$ is independent of any event of the form $Y \in B$. ($P\{X = i, Y = j\} = p_X(i) p_Y(j)$)

Binomial distribution

A random variable X is said to have the **Bernoulli distribution** with parameter p, where $0 \leq p \leq 1$, if $p_X(1) = p$ and $p_X(0) = 1 - p$. $E[X] = p$, $Var(X) = E[X^2] - E[X]^2 = p(1 - p)$

Suppose n independent *Bernoulli trials* are conducted, each resulting in a one with probability p and a zero with probability $1 - p$. Let X denote the total number of ones occurring in the n trials. The pmf of X is

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ for } 0 \leq k \leq n$$

$E[X] = np$ and $Var[X] = np(1 - p)$

Geometric distribution

Do Bernoulli trials untill the outcome of a trial is one. L denote the number of trials conducted. The pmf of L is:

$$p_L(k) = (1 - p)^{k-1} p \text{ for } k \geq 1$$

and $P\{L > k\} = (1 - p)^k$ for $k \geq 0$.
$E[L] = \dfrac{1}{p}$, $Var[L] = \dfrac{1 - p}{p^2}$

Negative binomial distribution

Let S_r denotes the number of trials required for r ones, and the last trail must be one. Let $n \geq r$, and let $k = n - r$. The event $\{S_r = n\}$ is determined by the outcomes of the ffirst n trials. The event is true iff there are $r - 1$ ones and k zeros in the first $k + r - 1$ trials, and trail n is one. Therefore, the pmf of S_r is given by

$$p(n) = \binom{n-1}{r-1} p^r (1 - p)^{n-r} \text{ for } n \geq r$$

$E[S_r] = \dfrac{r}{p}$, $Var(S_r) = rVar(L_1) = \dfrac{r(1 - p)}{p^2}$

Poisson disttribution

The **Poission probability distribution** with parameter $\lambda > 0$ is the one with pmf $p(k) = \dfrac{e^{-\lambda} \lambda^k}{k!}$ for $k \geq 0$. It's a good approximation for a binomial distribution with parameters n and p, when n is very large, p is very small, and $\lambda = np$.

$E[Y] = Var[Y] = \lambda$

Examples:

- *Radio active emissions in a fixed time interval*: n is the number of uranium atoms in a rock sample, and p is the probability that any particular one of those atoms emits a particle in a one minute period.
- *Incoming phone calls in a fixed time interval*: n is the number of people with cell phones within the access region of one base station, and p is the probability that a given such person will make a call within the next minute.
- *Misspelled words in a document*: n is the number of words in a document and p is the probability that a given word is misspelled.

Maximum likelihood parameter estimation

For a random variable X, and that the pmf of X is p_θ, where θ is a parameter. The probability of k being the observed value for X. The *likelihood* $X = k$ is $p_\theta(k)$. The *maximum likelihood estimate* of θ for observation k, denoted by $\hat{\theta}_{ML}(k)$, is the value of θ that maximizes the likelihood, $p_\theta(k)$, with respect to θ. (Give k, find θ (or p in some Bernoulli trials) to make $p_\theta(k)$ biggest, $\hat{\theta}_{ML}(k) =$the value of θ).

Markov and Chebychev inequalities and confidence intervals

Markov's inequality: $P\{Y \geq c\} \leq \dfrac{E[Y]}{c}$
Chebychev inequality: $P\{|X - \mu| \geq d\} \leq \dfrac{\sigma^2}{d^2}$

The law of total probability

$P(E_i|A) = \dfrac{P(AE_i)}{A} = \dfrac{P(A|E_i)P(E_i)}{P(A)}$
$E[X] = \sum_{j=1}^{J} E[X|E_j] P(E_j)$

confidence intervals

$$P\{p \in (\hat{p} - \frac{a}{2\sqrt{n}}, \hat{p} + \frac{a}{2\sqrt{n}})\} \geq 1 - \frac{1}{a^2}$$