The Usual Suspects: frequently occuring probability distributions

Robin Chapman

30 April 2008

I give information on the more frequently occurring probability distributions. For more details (and more distributions) see `http://en.wikipedia.org/wiki/Probability_distribution`.

1 Discrete distributions

In each case I give the probability that a variable with the given distribution takes each admissible value. Also I list mean, variance and usually the probability generating function. Recall this is

$$P_X(s) = E(s^X) = \sum_k P(X = k)s^k.$$

Also

$$P_X(1) = 1, \qquad P'_X(1) = E(X), \qquad P''_X(1) = E(X^2) - E(X).$$

1.1 The binomial distribution

Let $n \in \mathbf{N}$ and $0 < p < 1$. The *binomial distribution* with parameters n and p has range space $\{0, 1, \ldots, n\}$ and satisfies

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

We call this the $B(n, p)$ distribution. Then

$$E(X) = np, \qquad \mathrm{Var}(X) = np(1 - p), \qquad P_X(s) = (1 - p + ps)^n.$$

The binomial distribution arises when performing n independent trials each with probability p of success. The number of successes is a $B(n, p)$ random variable. The special case with $n = 1$ is sometimes called a *Bernoulli* random variable.

1.2 The geometric distribution

Let $0 < p < 1$. The *geometric distribution* with parameter p has range space $\mathbf{N} = \{1, 2, 3, \ldots\}$ and satisfies

$$P(X = k) = p(1-p)^{k-1}.$$

Then

$$E(X) = \frac{1}{p}, \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}, \qquad P_X(s) = \frac{ps}{1-(1-p)s}.$$

The geometric distribution arises when performing a sequence of independent trials each with probability p of success. The number of trials required to secure the first success is a geometric random variable with parameter p.

The geometric distribution is “memoryless”:

$$P(X > a + b \mid X > b) = P(X > a)$$

for each a, $b \in \mathbf{N}$.

1.3 The negative binomial distribution

Let $n \in \mathbf{N}$ and $0 < p < 1$. The *negative binomial distribution* with parameters n and p has range space $\mathbf{N} = \{n, n+1, n+2, \ldots\}$ and satisfies

$$P(X = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}.$$

Then

$$E(X) = \frac{n}{p}, \qquad \mathrm{Var}(X) = \frac{n(1-p)}{p^2}, \qquad P_X(s) = \frac{p^n s^n}{(1-(1-p)s)^n}.$$

The negative binomial distribution arises when performing a sequence of independent trials each with probability p of success. The number of trials required to secure the n-th success is a negative binomial random variable with parameters n and p.

When $n = 1$ the negative binomial distribution becomes the geometric distribution.

1.4 The hypergeometric distribution

Let $N \in \mathbf{N}$, $0 \leq n \leq N$ and $0 \leq a \leq N$. The *hypergeometric distribution* with parameters N, n and a satisfies

$$P(X = k) = \frac{\binom{a}{k}\binom{N-a}{n-k}}{\binom{N}{n}}.$$

We sometimes call this the $B(n, p)$ distribution. Then

$$E(X) = \frac{na}{N}, \qquad \mathrm{Var}(X) = \frac{na(N-a)(N-n)}{N^2(N-1)}$$

(I didn't compute these in lectures) while the generating function does not have a nice closed form.

The hypergeometric distribution arises when one takes a random n-element subset of an N-element set, of which a elements have some property A. The number of elements in the random n-element sets with property A is a hypergeometric random variable with parameters N, n and a.

1.5 The Poisson distribution

Let $\lambda > 0$. The *Poisson distribution* with parameter λ has range space $\mathbf{N}_0 = \{0, 1, 2, \ldots\}$ and satisfies

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

Then

$$E(X) = \lambda, \qquad \mathrm{Var}(X) = \lambda, \qquad P_X(s) = e^{\lambda(s-1)}.$$

When n is large and p is small the binomial distribution $B(n, p)$ is approximately Poisson with parameter np.

2 Continuous distributions

In each case I give the density function of a random variable with the given distribution. Also I list mean, variance and the moment generating function. Recall this is

$$M_X(t) = E(e^{tX}) = \sum_{m=0}^{\infty} E(X^m)\frac{t^m}{m!}.$$

2.1 The uniform distribution

Let $a < b$. The *uniform distribution* on the interval (a, b) has range space (a, b) and density function the constant

$$f_X(x) = \frac{1}{b-a}$$

on (a, b). Then

$$E(X) = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}, \qquad M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}.$$

2.2 The exponential distribution

Let $\lambda > 0$. The *exponential distribution* with parameter λ has range space $(0, \infty)$ and density function

$$f_X(x) = \lambda e^{-\lambda x}$$

for $x > 0$. Then

$$E(X) = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}, \qquad M_X(t) = \frac{\lambda}{\lambda - t} = \frac{1}{1 - t/\lambda}.$$

The exponential distribution is used for modelling waiting problems in continuous time, for instance decay of radioactive atoms, time for customers to join a queue, etc.

The exponential distribution is "memoryless":

$$P(X > a + b \mid X > b) = P(X > a)$$

for each $a > 0$, $b > 0$.

2.3 The gamma distribution

Let $n \in \mathbf{N}$ and $\lambda > 0$. The *gamma distribution* with parameters n and λ has range space $(0, \infty)$ and density function

$$f_X(x) = \lambda^n \frac{x^{n-1}}{(n-1)!} e^{-\lambda x}$$

for $x > 0$. Then

$$E(X) = \frac{n}{\lambda}, \qquad \mathrm{Var}(X) = \frac{n}{\lambda^2}, \qquad M_X(t) = \frac{\lambda^n}{(\lambda - t)^n} = \frac{1}{(1 - t/\lambda)^n}.$$

A more general version of the gamma distribution replaces the integer parameter n by a real parameter $\alpha > 0$. Then

$$f_X(x) = \lambda^\alpha \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}.$$

Here the gamma function is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx.$$

Then

$$E(X) = \frac{\alpha}{\lambda}, \qquad \mathrm{Var}(X) = \frac{\alpha}{\lambda^2}, \qquad M_X(t) = \frac{\lambda^\alpha}{(\lambda - t)^\alpha} = \frac{1}{(1 - t/\lambda)^\alpha}.$$

2.4 The Cauchy distribution

The *Cauchy distribution* has range space the whole real line **R** and density function

$$f_X(x) = \frac{1}{\pi(1 + x^2)}.$$

Since the integrals

$$\int_{-\infty}^\infty \frac{x\, dx}{1 + x^2} \qquad \text{and} \qquad \int_{-\infty}^\infty \frac{x^2\, dx}{1 + x^2}$$

are not absolutely convergent, the Cauchy distribution has no mean and no variance. It does not obey many of the usual laws of probability, for instance the Central Limit Theorem, since they in general apply only to random variables with well-defined mean and finite variance.

2.5 The normal distribution

Let $\mu \in \mathbf{R}$ and $\sigma > 0$. Then *normal distribution* with parameters μ and σ has range space **R** and density function

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}.$$

We call this the $N(\mu, \sigma^2)$ distribution. Then

$$E(X) = \mu, \qquad \mathrm{Var}(X) = \sigma^2, \qquad M_X(t) = e^{\mu t + \sigma^2 t^2/2}.$$

An $N(0,1)$ random variable is called a *standard* normal variable. If X is a standard normal variable then $Y = \sigma X + \mu$ is an $N(\mu, \sigma^2)$ random variable. Conversely if Y an $N(\mu, \sigma^2)$ random variable then $X = (X - \mu)/\sigma$ is a standard normal variable. The density function of a standard normal variable is denoted by ϕ:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

and its indefinite integral by Φ:

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-x^2/2}\, dx.$$

Certain random variables, for instance the binomial and Poisson, are approximately normal for certain (large) parameter values. The most important theorem in probability is the Central Limit Theorem. Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables each with the same distribution, well-defined mean μ and finite variance σ^2. Then for each u,

$$\lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} < u\right) = \Phi(u).$$

The normal distribution is used in many probability models. For instance the weights of manufactured objects, for example chocolate bars, are distributed approximately normally.