Random Variables and Probability Distributions – Prof. Richard B. Goldstein

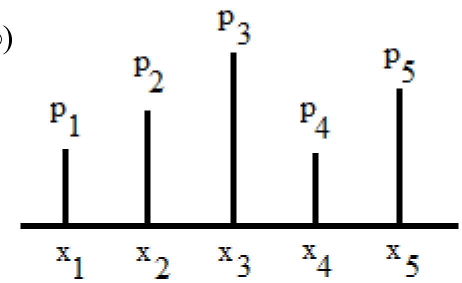
Random Variable – a variable which takes on values that are based upon the outcome of a random experiment

Discrete Distributions (takes on only a countable number of different values)

$f(x_i) = P\{X = x_i\} = p_i$ where $0 \le p_i \le 1$ and $\sum_{i=1}^{k} p_i = 1$ (k may be ∞)

Expected value $\mu = E[X] = \sum_{i=1}^{k} x_i p_i$

Variance $= \sigma^2 = Var[X] = \sum_{i=1}^{k} (x_i - \mu)^2 p_i = \sum_{i=1}^{k} x_i^2 p_i - \mu^2$



Continuous Distributions (takes on an infinite number of values over one or more intervals)

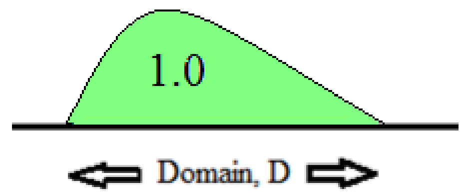
$f(x) \ge 0$ with an area under the curve is 1

Calculus Version (<u>optional</u>)

$\int_D f(x)\, dx = 1$ where D is the domain

$\mu = \int_D x f(x)\, dx$

$\sigma^2 = \int_D (x-\mu)^2 f(x)\, dx = \int_D x^2 f(x)\, dx - \mu^2$



Approximations

Binomial → Normal

As $n \to \infty$ one can approximate the binomial distribution by the normal distribution where $\mu = np$, $\sigma^2 = npq$ and corrections of ±0.5 are made when going from a discrete to a continuous distribution. The rule of thumb is both $np > 5$ and $nq > 5$. The approximation is best when $np \approx 0.5$.

Binomial → Poisson

If p is small and n is large the Poisson distribution can be used to approximate the binomial distribution. The smaller the p and larger the n, the better will be the approximation. The rule of thumb is $n \ge 100$ and $p < 0.1$.

Prof. Richard B. Goldstein – some discrete distributions using Excel

Binomial

BINOMDIST(x,n,p,c) = $\dfrac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$ if c = 0

x = number of successes
n = number of trials
p = probability of success on each trial
c = $\begin{cases} 0 \text{ for probability of } x \text{ successes} \\ 1 \text{ for cumulative probability} \end{cases}$

Examples

BINOMDIST(3,10,0.4,0) = 0.214991 P{3 successes out of 10 trials with p = 0.4}
BINOMDIST(3,10,0.4,1) = 0.382281 P{3 or fewer successes out of 10 trials with p = 0.4}

Hypergeometric

HYPGEOMDIST(x,n,M,N) = $\dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \dfrac{C_{M,x}C_{N-M,n-x}}{C_{N,n}}$

x = successes in the sample
n = sample size
M = successes in the population
N = population size

Example

Five cards are drawn from a deck of 52 playing cards. This formula calculates the probability that two of the five cards are hearts:

HYPGEOMDIST(2,5,13,52) = 0.27428

Poisson

POISSON(n,λ,c) = $\dfrac{e^{-\lambda}\lambda^n}{n!}$

n = number of events
λ = expected numeric value for the mean of the distribution
c = $\begin{cases} 0 \text{ for probability of } n \text{ events} \\ 1 \text{ for cumulative probability of } 0 \text{ to } n \text{ events} \end{cases}$

Example

In a typical hour 30 customers arrive in a bank. What is the probability that 35 customers arrive?

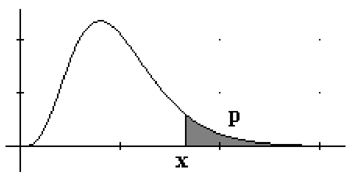
POISSON(35,30,0) = 0.045308

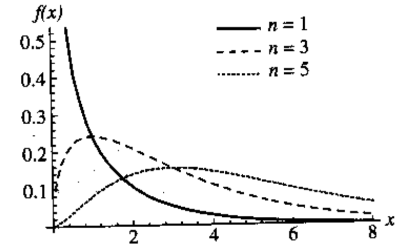
Prof. Richard B. Goldstein – some continuous distributions using Excel

Chi-Square

$$\text{CHIDIST}(x, n) = \int_0^{\infty} \frac{1}{2^{n/2}\Gamma(n/2)} t^{n/2-1} e^{-t/2} dt$$

x = independent variable
n = number of degrees of freedom





Chi-Square Densities

CHIDIST(36.41503,24) = 0.05 (the value p)
CHIINV(p, n) = CHIINV(0.05,24) = 36.41503 (the value x)

Exponential

EXPONDIST(x, λ, c) = $\lambda e^{-\lambda x}$ if c = 0 and cumulative probability if C =1

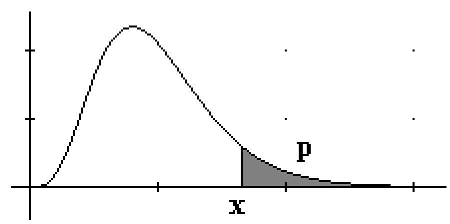
x = independent variable
λ = parameter = 1/mean

F

$$\text{FDIST}(X, N_1, N_2) = \frac{N_1^{N_1/2} N_2^{N_2/2}}{\beta(N_1, N_2)} \int_0^{x} t^{(N_1-2)/2} (N_2 + N_1 t)^{-(N_1+N_2)/2} dt$$

X = independent variable
N_1 = numerator degrees of freedom
N_2 = denominator degrees of freedom

FDIST(6.256057,5,4) = 0.05 (the value p)
FINV(0.05,5,4) = 6.256057 (the value x)

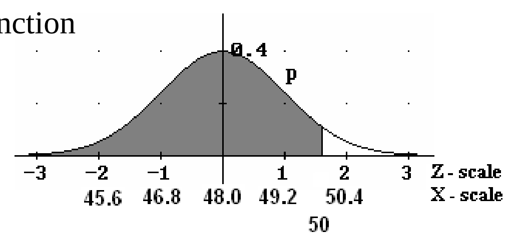


Normal

$$\text{NORMDIST}(x, \mu, \sigma, c) = \int_{-\infty}^{x} \frac{e^{-(t-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} dt$$

x = value at which to evaluate function
μ = mean of the normal distribution
σ = standard deviation of the normal distribution
c = 1 to return the cumulative normal distribution function
0 (the default) to return the probability density function

NORMDIST(50,48,1.2,1) = 0.95221 (cumulative prob. p)
NORMDIST(50,48,1.2,0) = 0.082898 (density value)
NORMINV(p,μ,σ) = NORMINV(0.95221,48,1.2) = 50 (x)

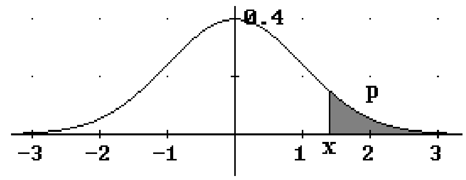


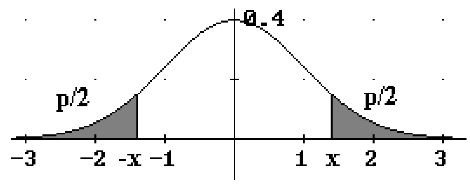
Prof. Richard B. Goldstein – some continuous distributions using Excel

Student t

$$\text{TDIST}(x, n, \text{tails}) = 1 - \frac{1}{\sqrt{n}\,\beta(\frac{1}{2}, \frac{n}{2})} \int_{-x}^{x} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt \quad \text{for 2 tails}$$

x = Value at which to evaluate the distribution.
n = Integer number of degrees of freedom
tails = 1 to return the area in a one-tailed distribution
2 to return the area in a two-tailed distribution





TDIST(2.228139,10,2) = 0.05 (area in each tail is 0.025)
TDIST(p, df) = TINV(0.05,10) = 2.228 (for 2 tails only)

Other Distributions :

Beta	BETADIST	BETAINV
Gamma	GAMMADIST	GAMMAINV
Log-Normal	LOGNORMDIST	no inverse function
Weibull	WEIBULL	no inverse function