

2.13 Probability and statistics

Discrete statistics

Mean	$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i$	(2.541)	x_i	data series
			N	series length
			$\langle \cdot \rangle$	mean value
Variance ^a	$\text{var}[x] = \frac{1}{N-1} \sum_{i=1}^N (x_i - \langle x \rangle)^2$	(2.542)	$\text{var}[\cdot]$	unbiased variance
Standard deviation	$\sigma[x] = (\text{var}[x])^{1/2}$	(2.543)	σ	standard deviation
Skewness	$\text{skew}[x] = \frac{N}{(N-1)(N-2)} \sum_{i=1}^N \left(\frac{x_i - \langle x \rangle}{\sigma} \right)^3$	(2.544)		
Kurtosis	$\text{kurt}[x] \simeq \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \langle x \rangle}{\sigma} \right)^4 \right] - 3$	(2.545)		
Correlation coefficient ^b	$r = \frac{\sum_{i=1}^N (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sqrt{\sum_{i=1}^N (x_i - \langle x \rangle)^2} \sqrt{\sum_{i=1}^N (y_i - \langle y \rangle)^2}}$	(2.546)	x, y	data series to correlate
			r	correlation coefficient

^aIf $\langle x \rangle$ is derived from the data, $\{x_i\}$, the relation is as shown. If $\langle x \rangle$ is known independently, then an unbiased estimate is obtained by dividing the right-hand side by N rather than $N-1$.

^bAlso known as “Pearson’s r .”

Discrete probability distributions

distribution	$\text{pr}(x)$	mean	variance	domain		
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$	$(x=0, 1, \dots, n)$	(2.547)	$\binom{n}{x}$ binomial coefficient
Geometric	$(1-p)^{x-1} p$	$1/p$	$(1-p)/p^2$	$(x=1, 2, 3, \dots)$	(2.548)	
Poisson	$\lambda^x \exp(-\lambda)/x!$	λ	λ	$(x=0, 1, 2, \dots)$	(2.549)	



Continuous probability distributions

distribution	pr(x)	mean	variance	domain	
Uniform	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$(a \leq x \leq b)$	(2.550)
Exponential	$\lambda \exp(-\lambda x)$	$1/\lambda$	$1/\lambda^2$	$(x \geq 0)$	(2.551)
Normal/ Gaussian	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$	μ	σ^2	$(-\infty < x < \infty)$	(2.552)
Chi-squared ^a	$\frac{e^{-x/2} x^{(r/2)-1}}{2^{r/2} \Gamma(r/2)}$	r	$2r$	$(x \geq 0)$	(2.553)
Rayleigh	$\frac{x}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right)$	$\sigma\sqrt{\pi/2}$	$2\sigma^2\left(1 - \frac{\pi}{4}\right)$	$(x \geq 0)$	(2.554)
Cauchy/ Lorentzian	$\frac{a}{\pi(a^2 + x^2)}$	(none)	(none)	$(-\infty < x < \infty)$	(2.555)

^aWith r degrees of freedom. Γ is the gamma function.

Multivariate normal distribution

Density function	$\text{pr}(\mathbf{x}) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})^T\right]}{(2\pi)^{k/2}[\det(\mathbf{C})]^{1/2}}$	(2.556)	pr probability density k number of dimensions $\mathbf{C}$ covariance matrix $\mathbf{x}$ variable (k dimensional) $\boldsymbol{\mu}$ vector of means T transpose det determinant μ_i mean of i th variable σ_{ij} components of $\mathbf{C}$ r correlation coefficient x_i normally distributed deviates y_i deviates distributed uniformly between 0 and 1
Mean	$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_k)$	(2.557)	
Covariance	$\mathbf{C} = \sigma_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$	(2.558)	
Correlation coefficient	$r = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$	(2.559)	
Box-Muller transformation	$x_1 = (-2 \ln y_1)^{1/2} \cos 2\pi y_2$ $x_2 = (-2 \ln y_1)^{1/2} \sin 2\pi y_2$	(2.560) (2.561)	



Random walk

One-dimensional	$\text{pr}(x) = \frac{1}{(2\pi Nl^2)^{1/2}} \exp\left(\frac{-x^2}{2Nl^2}\right) \quad (2.562)$	x displacement after N steps (can be positive or negative) $\text{pr}(x)$ probability density of x ($\int_{-\infty}^{\infty} \text{pr}(x) dx = 1$) N number of steps l step length (all equal)
rms displacement	$x_{\text{rms}} = N^{1/2}l \quad (2.563)$	x_{rms} root-mean-squared displacement from start point
Three-dimensional	$\text{pr}(r) = \left(\frac{a}{\pi^{1/2}}\right)^3 \exp(-a^2r^2) \quad (2.564)$ where $a = \left(\frac{3}{2Nl^2}\right)^{1/2}$	r radial distance from start point $\text{pr}(r)$ probability density of r ($\int_0^{\infty} 4\pi r^2 \text{pr}(r) dr = 1$) a (most probable distance) ⁻¹
Mean distance	$\langle r \rangle = \left(\frac{8}{3\pi}\right)^{1/2} N^{1/2}l \quad (2.565)$	$\langle r \rangle$ mean distance from start point
rms distance	$r_{\text{rms}} = N^{1/2}l \quad (2.566)$	r_{rms} root-mean-squared distance from start point

Bayesian inference

Conditional probability	$\text{pr}(x) = \int \text{pr}(x y') \text{pr}(y') dy' \quad (2.567)$	$\text{pr}(x)$ probability (density) of x $\text{pr}(x y')$ conditional probability of x given y'
Joint probability	$\text{pr}(x, y) = \text{pr}(x) \text{pr}(y x) \quad (2.568)$	$\text{pr}(x, y)$ joint probability of x and y
Bayes' theorem ^a	$\text{pr}(y x) = \frac{\text{pr}(x y) \text{pr}(y)}{\text{pr}(x)} \quad (2.569)$	

^aIn this expression, $\text{pr}(y|x)$ is known as the posterior probability, $\text{pr}(x|y)$ the likelihood, and $\text{pr}(y)$ the prior probability.

