Distributions

Discrete

`Combinatorial` Finite space, equally likely outcomes
$P(A) = card(A)/n$

`Bernoulli` Success probability of bent coin
$p_0 = q = (1-p)$ and $p_1 = p$
$\mu = p, \sigma^2 = pq$

`Binomial` Probability of n heads in k tosses
$p_k = \binom{n}{k} p^k (1-p)^{n-k}$
$\mu = np, \sigma^2 = npq$

`Geometric` Failures before 1st success, $k \geq 0$
$p_k = (1-p)^k p$
$\mu = \frac{q}{p}, \sigma^2 = \frac{q}{p^2}$

`Negative Binomial` Failures before rth success $k \geq 0$
$w_r(k;p) = \binom{k+r-1}{k} q^k p^r$
Note the negative binomial is sum of r **independent** geometric RVs.
$\mu = \frac{rq}{p}, \sigma^2 = \frac{rq}{(p)^2}$

`Poisson` Characterization of rare events
$p_k = e^{-\lambda}\lambda^k/k!$
$\mu = \lambda, \sigma^2 = \lambda$

Continuous

`Uniform` Agnostic about outcome on interval
$f(x) = 1/(b-a)$ on
$a < x < b$ and 0 $o.w.$
$\mu = \frac{1}{2}(a+b), \sigma^2 = \frac{1}{12}(b-a)$

`Gamma` Note $\Gamma = (n-1)!$
$g_n(t;\alpha) = \alpha \frac{(\alpha t)^{n-1}}{(n-1)!} e^{-\alpha t}$
$\mu = \frac{1}{\alpha}, \sigma^2 = \frac{1}{\alpha^2}$

`Chi-squared` Finite X_k $\mathcal{N}(0,1)$; $V^2 = \sum_{k=1}^n X_k^2$
$g_{n/2}(t;1/2) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} t^{\frac{n}{2}-1} e^{-\frac{t}{2}}$
$\mu = n, \sigma^2 = 2n$

`Exponential` A model for true randomness
$f(x) = \alpha e^{-\alpha x}$ for $x > 0$
$\mu = \frac{1}{\alpha}, \sigma^2 = \frac{1}{\alpha^2}$

`Normal` The one to rule them all...
$f(x) = \frac{1}{\sqrt{(2\pi)}\sigma} e^{-(x-\mu)^2/2\sigma^2}$

Even function, strictly positive, dies monotonically...
Can be sometimes written as $N(\mu, \sigma^2)$...

`Multivariate Normal` Note that x and m are vectors

$$f(x) = \frac{1}{2\pi^{\frac{n}{2}} det(A)^{\frac{1}{2}}} exp(-\frac{1}{2}(x-m)A^{-1}(x-m)^T)$$

Negative Binomial Trick: $\binom{r+k-1}{k} q^k p^r = \binom{-r}{k}(-q)^k p^r$

Properties of Distributions

Shifting and Scaling *If two densities come from same type:* $\frac{1}{a} f(\frac{x-m}{a})$ *then* $\mu = a\mu + m$ $\sigma^2 = a^2\sigma^2$
Convolution *The convolution of any pair of distributions must necessarily be a distribution*
Stability of the Normal *Suppose* $X_1, X_2, ...X_n$ *are independent and for each k* X_k *is normal with mean* μ_k *and variance* σ_k. *Then* $S_n = \sum_{k=1}^n X_k$ *is normal with mean* $\sum_{k=1}^n \mu_k$ *and variance* $\sum_{k=1}^n \sigma_k$, *or in other words normal densities are stable under convolution. Also stable under* ***non-degenerate*** *linear transformation*
Translation Invariance: Uniform
Sum of Independent Poisson RVs *For* ***independent*** *random variables drawn from Poisson distribution is Poisson and the mean is* $\lambda_1 + \lambda_2$
Sum of Independent Binomials is also binomial.
Sum of Independent Gamma RVs *Given* $X_1 \sim g_{n_1}(x_1, \alpha)$ *and* $X_2 \sim g_{n_2}(x_2, \alpha)$ *then* $Z = X_1 + X_2 \sim g_{n_1+n_2}(z, \alpha)$. *Note the RV's must have the same mean* $\frac{1}{\alpha}$
The marginals of a multivariate normal are normal *Note: a system of variables with marginal normal densities need not have a jointly normal density, and, in fact a joint density need not even exist*
Sum of Exponentials *Sum of* ***independent*** *exponential random variables from a common exponential density* $\alpha e^{-\alpha x}$ *for* $x > 0$ *is* S_n, *the sum* $S_n = X_1 + X_2 + ... + X_n$ *has the* ***gamma density***.
Memoryless Property ***ONLY*** *the exponential and geometric distributions are memoryless. The random variable X exhibits the memoryless property if* $P\{X > s+t | X > s\} = P\{X > t\}$ *or recast* $P\{X > s+t|X > s\} = P\{X > t\}$ *if and only if* $P\{X > s+t\} = P\{X > s\}P\{X > t\}$
Rotation of Bivariate Normal *There exists a rotation of a bivariate normal st. the normal coordinates become independent*
Max variance for Bernoulli Trial *Is 1/4 which occurs if* $p = 1/2$. *Bernoulli trials are assumed to be independent unless otherwise stated.*
Limit of the geometric *The exponential distribution is the continuous limit of the geometric distribution. Use two integration by parts to show variance is* $\frac{1}{\alpha^2}$ *and the mean is* $\frac{1}{\alpha}$
The Gamma Density is Closed Under Convolutions *For every fixed* $\alpha > 0$ *the family of gamma densities* $\{g_v(;\alpha), v > 0\}$ *is closed under convolutions, that is,* $(g_\mu \star g_\nu)(x) = g_{\mu+\nu}(x)\forall \mu\nu$. *Restatement of earlier property.*

Expectation

Variance $Var(X) = E(X^2) - E(X)^2$
Covariance $Cov(X,Y) = E(XY) - E(X)E(Y)$
Law of the Unconcious Statistician *If X is a discrete r.v. and g is a function from* $R \to R$ *then,* $E(g(X)) = \sum_x g(x)P(X=x)$
Independence *Independent variables are uncorrelated, but uncorrelated variables are not necessarily independent. However, this is a necessary and sufficient condition if we are talking about the marginals of a normal density* **Additivity** *Expectation is additive. Variance is additive if the summands are independent.*
Conditional Expectation $E(X_2) = E(E(X_2|X_1))$

Conditional Probability

$P(A|H) = \frac{P(A \cap H)}{P(H)}$
$P_H : A \to P(A|H)$
$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1|A_2 \cap ... \cap A_n)$
$\times P(A_2|A_3 \cap ... \cap A_n) \times P(A_{n-1}|P(A_n) \times P(A_n)$
Bayes Rule $P(A_k|H) = \frac{P(H|A_k)P(A_k)}{\sum_{j \geq 1} P(H|A_j)P(A_j)}$

Inclusion-Exclusion

$S_k = \sum(P(A_{j_1}) \cap ... \cap P(A_{j_k})) = \binom{n+1}{k}(1-ka)_+^n$
De Finneti's Theorem
$P\{L_1 > x_1, ..., L_{n=1} > x_{n+1}\} = (1 - \frac{x_1}{\tau} - \frac{x_2}{\tau} ... - \frac{x_{n+1}}{\tau})_+^n$

Limit Laws

Weak Law of Large Numbers(Khinchin)
If $X_1, X-2, ...$ is a sequence of independent random variable with common distibution with finite mean μ and let $S_n = (X_1 + ... + X_n)$ for each n. Then $\frac{1}{n}S_n \to^p \mu$. Or, $P\{|\frac{1}{n}S_n - \mu| \geq \epsilon\} \to 0$ as $n \to \infty$ for every choice $\epsilon > 0$
Strong Law of Large Numbers(Kolmogorov, conditions same) $\frac{1}{n}S_n \to^{a.e.} \mu$. Or, $P\{|\frac{1}{n}S_n - \mu| \geq \epsilon \; i.o.\} = 0$ for every $\epsilon > 0$.
Central Limit Theorem
Suppose $X_1, X-2, ...$ *is a sequence of independent random variable drawn from a common distribution F with mean zero and variance one. For each n, let* $S_n* = (X_1 + ... + X_n)/\sqrt{n}$. *Then* S_n* *converges in distribution to the standard normal.* $P\{a < S_n* < b\} \to \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx$ *If* $E(X_k) = \mu$ *and* $Var(X_k) = \sigma^2$ *then we can center and scale the variables such that* $S_n* = (S_n - n\mu/\sqrt{n}\sigma)$ *and the theorem holds for the new properly normalize* S_n*.

Inequalities

Convexity The chord lies above the curve
$\Psi(\alpha x + (1-\alpha)y) \leq \alpha\Psi(x) + (1-\alpha)\Psi(y)$
Jensen If Ψ is convex , X is integrable, and $\Psi(x)$ is integrable then: $\Psi(E(x)) \leq E(\Psi(x))$
Another of Jensen $E(X^2) \geq E(X)^2$
AM-GM $x_1^{p_1} x_2^{p_2} ... x_n^{p_n} \leq p_1 x_1 + p_2 x_2 + ... p_n x_n$
Specialization of AM-GM $x^{1/p} y^{1/q} \leq (1/p)x + (1/q)y$ simplifies to: $xy \leq (1/p)x^p + (1/q)y^p$
Holder's $|E(XY)| \leq E(|XY|) \leq E(|X|)^{1/p} E(|Y|)^{1/q}$
Cauchy-Schwarz If $p = q = 2$ in Holder's $\to$ $|E(XY)^2| \leq E(X^2)E(Y^2)$
Minkowski $p \geq 1$ then $||x+y||_p \leq ||x||_p + ||y||_p$
Chebyshev X has finite second moment and t is strictly positive. Then, $P\{|X - E(X)| \geq t\} \leq Var(X)/t^2$
If $X_1, X-2, ...$ is a sequence of independent random variable with common distibution F. For each n, let $S_n = (X_1 + ... + X_n)$.
$P\{|S_n - n\mu| \geq n\epsilon\} \leq \frac{n\sigma^2}{n^2\epsilon^2}$
Chernoff $P\{S_n \geq t\} \leq (\inf_{\lambda \geq 0} e^{-\lambda t/n} M(\lambda))^n$ and
$P\{S_n \leq t\} \leq (\inf_{\lambda \geq 0} e^{-\lambda t/n} M(-\lambda))^n$
Where $M(\lambda) = E(e^{\lambda X_j}) = \int_{\mathcal{R}} e^{\lambda x} dF(x)$ is the moment generating function
Markov $P\{X \geq a\} \leq \frac{E(X)}{a}$
Normal Tail Bound p 317 $\frac{\phi(x)}{x}(1 - \frac{1}{x^2}) < 1 - \Phi(x) < \frac{\phi(x)}{x}$
Normal Tail Bound p 165 $\int_t^\infty \phi(x)dx \leq \frac{1}{2}e^{-t^2/2}$

Logarithmic Identities

$y = \log_b(x)$ iff $x = b^y$
$\log_b\left(\frac{x}{y}\right) = \log_b(x) - \log_b(y)$
$\log_b(x) = \log_b(c)\log_c(x) = \frac{\log_c(x)}{\log_c(b)}$
$\log_b(x^n) = n\log_b(x)$
$\log_b(1) = 0$
$\log_b(b) = 1$
$\log_b(xy) = \log_b(x) + \log_b(y)$
$log(\infty) \to \infty$ $log(0) \to -\infty$ $log(1) = 0$

Exponential Identities

$x^a x^b = x^{(a+b)}$ $(x^a)^b = x^{(ab)}$
$x^a y^a = (xy)^a$ $x^{(a-b)} = \frac{x^a}{x^b}$

Common Infinite Series

Exponential

$\sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z$

$\sum_{k=0}^{\infty} k\frac{z^k}{k!} = ze^z$ Mean of the Poisson

$\sum_{k=0}^{\infty} k^2\frac{z^k}{k!} = (z+z^2)e^z$ Second moment of Poisson

$\sum_{k=0}^{\infty} k^3\frac{z^k}{k!} = (z+3z^2+z^3)e^z$

$\sum_{k=0}^{\infty} k^4\frac{z^k}{k!} = (z+7z^2+6z^3+z^4)e^z$

Binomial

$\sum_{k=0}^{\infty} \binom{\alpha}{k} z^k = (1+z)^\alpha, |z|<1$

$\sum_{k=0}^{n} \binom{n}{k} x^{n-k}y^k = (x+y)^n$

Integration

$\int x^n dx = \frac{1}{n+1}x^{n+1},\ n \neq -1$

$\int \frac{1}{x}dx = \ln|x|$

$\int u\,dv = uv - \int vdu$

$\int e^x dx = e^x$

$\int a^x dx = \frac{1}{\ln a}a^x$

$\int \ln x dx = x\ln x - x$

$\int \sin x dx = -\cos x$

$\int \cos x dx = \sin x$

The Bernoulli Schema

Theorem 1 *Suppose $X_1, ..., X_n$ is a sequence of Bernoulli tials with success probability p. Then the sume $S_n = X_1 + ... + X_n$ has a distribution $b_n(k) = \binom{n}{k}p^k q^{n-k}$. Recall as a matter of convention the definition for real t and integer k that $\binom{t}{k} = t(t-1)(t-2)...(t-k+1)/k!$ if $k \geq 0$ and 0 otherwise.*

Example: Ball and Urn *Put n balls in m urns, probability that there are k balls in first r: $\binom{n}{k}\frac{r}{m}^k(1-\frac{r}{m})^{n-k}$*

Example: Polls *Select n individuals from population, an unknown fraction p support policy. All samples are independent.$S_n = \sum_{j=1}^{n} = Z_j$ is the number of individuals in favor, $S_n \sim b_n(k;p)$. Let $\zeta \in [0,1]$ be a guess for p. $b(\zeta) = b_n(S_n;\zeta) = \binom{n}{S_n}\zeta^{S_n}(1-\zeta)^{n-S_n}$, then $b(\zeta)$ is maximized at $\zeta = (S_n/n)$. The sample mean is $\arg max_\zeta b_n(S_n);\zeta)$ which is the mean of the empirical distribution. Thus, p estimated by this value yields the largest probability of observations consistent with the data. We say that the sample mean is the unbiased estimator of p*

Error Bound on Sample Mean *Via Chebyshev*

$P\{|\frac{1}{n}S_n - p| \geq \epsilon\} = \sum b_n(k;p) \leq \frac{pq}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$

Example: Random Walks *The number of paths from (a,b) to (a',b') is $N_n(k) = \binom{n}{\frac{n+k}{2}}$ where $n = a' - a$ and $k = b' - b$. If the path goes through strictly positive values the number of such paths is $N_n^+(k) = N_{n-1}(k-1) - N_{n-1}(k+1) = \binom{n-1}{\frac{n+k}{2}-1} - \binom{n-1}{\frac{n+k}{2}} = \frac{k}{n}\binom{n}{\frac{n+k}{2}}$*

Example: Returns *The probability of a return is*

$P\{S_2 = N_{2v}(0)2^{-2v} = \binom{2v}{v}2^{-2v}$

Example: Waiting Time *The waiting time is the wait until the first succes $w(k;p) = P(W=k) = q^k p(k \geq 0)$. Now $\sum_k w(k;p) = p\sum_{k=0}^{\infty} q^k = p/(1-q) = 1$. Thus $w(k;p)$ is called the geometric distribution. Useful manipulation:*

$\sum_k^{\infty} q^k = 1/(1-q)$

Sum of Independent RVs geometric

Suppose $\{W_i, i \geq 1\}$ is a sequence of independent geometric random variables with common success parameter $p > 0$. For each r, let $T_r = W_1 + W_2 + ... + W_r$ and let $w_r(k;p) = P(T_r = k)$ be its distribution. Then $w_r(k;p) = \binom{r+k-1}{k}q^k p^r$ this is the negative binomial.

Problem of the Points $P_m, n(p) = \sum_{k=0}^{n-1}\binom{-m}{k}(-q)^k p^m$ *where m is the require points to win and n is the maximum number of failures*

Example: Population Size *Hypergeometric Distribution:*

$h(i) = h_{m,n}(i;r) = \frac{\binom{m}{i}\binom{n-m}{r-i}}{\binom{n}{r}}$ *where $(0 \leq u \leq r)$*

Vandermonde's Convolution $\sum_i \binom{m}{i}\binom{n-m}{r-i} = \binom{n}{r}$ *can be used to show hypergeometric is a density*

Another funny sum $\binom{n}{0}^2 + \binom{n}{1}^2 + ... + \binom{n}{n}^2 = \binom{2n}{n}$

The Essence of Randomness

Buffon's Needle

Gamma Function etc *Continuous version*

$g_v(x,\alpha) = \frac{\alpha}{\Gamma(v)}(\alpha x)^{v-1}e^{-\alpha x}$

The Coda of the Normal

Bivariate Normal

Suppose (X_1, X_2) is a random pair governed by the bivariate normal density $\phi(x_1, x_2, \rho)$ and $Y = X_1 + X_2$ then X_1 and X_2 share a common marginal normal distribution with zero mean and unit variance. Moreover, $Cov(X_1, X_2) = E(X_1X_2) = \rho$ By additivity of expectation $E(Y) = E(X_1) + E(X_2) = 0$, while, $Var(Y) = E((X_1+X_2)^2) = E(X_1^2)+E(X_2^2)+2E(X_1X_2) = 2(1+\rho)$

Random

The Hockey Stick Identity $\sum_{j=k}^{n}\binom{j}{n} = \binom{n+1}{k+1}$

Convergence

A sequence of random variables $\{X_1, X_2, ...\}$ converges almost surely to a random variable X if

$P\{s \in S : lim_{n\to\infty} X_n(s) = X_s\} = 1$