Multiplication and permutation rules
sample k is drawn from a population of n distinc objects:
Order doesn't matter and replace: C_{n+k-1}^k
Order doesn't matter and no replace: C_n^k
Probability Axioms
$P(A\bigcup B\bigcup C) = P(A)+(B)+P(C)-P(A\bigcap B)-P(A\bigcap C)-P(B\bigcap C)+P(A\bigcap B\bigcap C)$
Conditional probability: $P(B|A) = P(A\bigcap B)/P(A)$
Independent events: two events are independent if: 1) $P(A|B) = P(A)$; 2) $P(B|A) = P(B)$; 3) $P(A\bigcap B) = P(A)P(B)$
Bayes theorem & Conditional probabilities
$P(A|B) = P(B|A)P(A)/P(B)$
Discrete Distribution *mean*: measure of center of mass; 1st moment; $\mu = E(X) = \sum_x x * P(X=x)$
variance: measure of dispersion; 2nd moment; $\sigma^2 = V(X) = \sum_x (x-\mu)^2 f(x) = E(x^2) - \mu^2$ (can be infinite: $P(X=x) \geq 1/x^3$)
skewness: how asymmetric is the distribution around the mean. Normalized 3-rd moment: $\gamma = E((x-\mu)^3/\sigma^3)$ (can be infinite: $P(X=x) \geq 1/x^4$)
geometirc mean: for very broad distribution. Mean is dominated by very unlikely but very large events (like lottery). It is $exp(E(logX))$.
NOTE: All can be infinite.
Discrete uniform distribution
$f(x) = 1/(b-a+1)$, a, b is integer. $\mu = (b+a)/2$, $\sigma^2 = [(b-a+1)^2-1]/12$
Bernouli distribution
$f(x) = p, if\ x = 1; 1-p, if\ x = 0$. $E(X) = p; Var(X) = p(1-p)$
Binomial distribution
Sum of n independent bernouli trials, $f(x) = C_x^n p^x (1-p)^{n-x}$. $E(X) = np; Var(X) = np(1-p)$
Poisson distribution
$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$. $E(X) = Var(X) = \lambda$

- covered genome fraction: $coverage = \lambda = NL/G; P(X > 0) = 1 - exp(-\lambda); G_{covered} = G * P(X > 0)$
- how many configs: modified $\lambda = (N-1)(L - L_{ov}/G)$, probability no left ends fall inside a read, $N_{config} = Nexp(-\lambda)$
- average length of config: $G_{covered}/N_{config}$

Geometric distribution
continue until success: $P(X=x) = p(1-p)^{x-1}$. $E(X) = 1/p; Var(X) = (1-p)/p^2$
Example: time to last common maternal ancestor: $P(T=t) = (1-1/N)^{t-1}(1/N)$
Negative binomial distribution number of trials until r successes: $f(x) = C_{r-1}^{x-1} p^r (1-p)^{x-r}$. $E(X) = r/p; Var(X) = r(1-p)/p^2$
Example: cancer passenger and driver mutation
Power Law Distribution
$P(X=x) = Cx^{-\lambda}$, where C is normlization term, $1 = \sum_x C.x^{-\lambda}$ -¿ $C = 1/\zeta(\lambda)$. Mean and variance can be infinite.
Example: protein-protein network; cancer mutation
Continuous Distribution
PDF is the derivative of CDF: $f(x) = \frac{dF(x)}{dx}$. $E(X) = \int_{-\infty}^{\infty} xf(x)dx; Var(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$
Contunous uniform distribution
$f(x) = 1/(b-a)$. $E(X) = (b+a)/2; Var(X) = (b-a)^2/12$
Constant rate (poisson process)
Discrete events happen at rate r; expected #events in time x is rx. The actual #events N_x is a poisson distribution discrete random variable. $p(N_x = n) = \frac{(rx)^n}{n!}exp(-rx)$. $E(N_x) = pL = rx$
Exponential Distribution
Models the time interval to the 1st event. Exponential random variable X describes **interval** between 2 successes of a constant rate random process with success rate r per unit interval.

- PDF: $f(x) = re^{-rx}, 0 \leq x < \infty$
- CCDF: $P_x(X > x) = P_N(N_x = 0) = exp(-rx)$
- $u = E(X) = \frac{1}{r}$ and $\sigma^2 = V(X) = \frac{1}{r^2}$

the only memoryless distribution: $P(x > t+s | x > s) = P(x > t)$
Erlang Distribution
Models the time interval to the k^{th} event, a sum of k exponentially distributed variables.
$P(X > x) = \sum_{m=0}^{k-1} \frac{e^{-rx}(rx)^m}{m!} = 1 - F(x)$
$f(x) = F(x)' = \frac{r^k x^{k-1} e^{-rx}}{(k-1)!}$
Gamma Distribution
random variable x with PDF as $f(x) = \frac{r^k x^{k-1} e^{-rx}}{\Gamma(k)}$ has a gamma random distribution. If k is an positive integer, X has an Erlang distribution.
$\int_0^\infty f(x)dx = 1 \rightarrow \Gamma(k) = \int_0^\infty r^k x^{k-1} e^{-rx} dx = \int_0^\infty y^{k-1} e^{-y} dy,\ where\ y = rx$
Properties of Gamma function:

- $\Gamma(1) = 1$
- $\Gamma(k) = (k-1)\Gamma(k-1)$, recursive property
- $\Gamma(k) = (k-1)!$, factorial function
- $\Gamma(1/2) = \pi^{1/2} = 1.77$

Mean and Variance of Erlang and Gamma: $\mu = E(X) = k/r$, $\sigma^2 = V(x) = k/r^2$
Normal/Gaussian Distribution
$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \sim N(\mu, \sigma)$ The sum of many independent random variables could be approximated with a Gaussian.
Standard Normal Distribution: $Z \sim N(0,1)$. CDF is $\Phi(z) = P(Z \leq z)$
$N \sim (\mu, \sigma)$ can be **standardized** into $N \sim (0,1)$ by $Z = \frac{X-\mu}{\sigma} \rightarrow P(X \leq x) = P(Z \leq z)$
Lognormal Distribution
$X = e^W$, where $W\ N(\theta, \omega) \rightarrow W = ln(X)$ X is a lognoraml distribution variable. $F(x) = P(X < x) = P(exp(W) \leq x) = P(W \leq ln(x)) = P(Z \leq \frac{ln(x)-\theta}{\omega}) = \Phi(\frac{ln(x)-\theta}{\omega})$ for $x > 0$; or 0 if $x \leq 0$
$f(x) = \frac{dF(x)}{dx} = \frac{1}{x}\frac{1}{\omega\sqrt{2\pi}} exp(-(\frac{ln(x)-\theta}{2\omega})^2)$ for $x > 0$
$E(X) = e^{\theta+\omega^2/2}$ and $V(X) = e^{2\theta+\omega^2}(e^{\omega^2}-1)$
Joint Probability Distribution
Joint PMF, $f_{XY}(x,y)$
Marginal probability distribution: 1) $f_X(x) = \sum_y f_{XY}(x,y)$; 2) $f_Y(y) = \sum_x f_{XY}(x,y)$

Use marginal distributions to compute E and V:

y = number of times city name is stated	x = number of bars of signal strength					
	1	2	3	$f(y)$ =	$y*f(y)$ =	$y^2*f(y)$ =
1	0.01	0.02	0.25	0.28	0.28	0.28
2	0.02	0.03	0.20	0.25	0.50	1.00
3	0.02	0.10	0.05	0.17	0.51	1.53
4	0.15	0.10	0.05	0.30	1.20	4.80
$f(x)$ =	0.20	0.25	0.55	1.00	2.49	7.61
$x*f(x)$ =	0.20	0.50	1.65	2.35		
$x^2*f(x)$ =	0.20	1.00	4.95	6.15		

$E(X) = 2.35; V(X) = 6.15 - 2.35^2$ $E(Y) = 2.49; V(X) = 7.61 - 2.49^2$

Conditional probability distribution: $P(Y = y|X = x) = P(X = x, Y = y)/P(X = x) = f(x, y)/f_X(x)$

Random variables independent if all events A that Y=y and B that X=x are independent if any one of the conditions is met:

- $P(Y = y|X = x) = P(Y = y)$
- $P(X = x|Y = y) = P(X = x)$
- $P(X = x, Y = y) = P(X = x)P(Y = y)$ for every pair of x and

Conditional probability density function: $f_{Y|x}(y) = \dfrac{f_{XY}(x, y)}{f_X(x)}$

Independence of continuous random variable:

- $f_{XY}(x, y) = f_X(x)f_Y(y)$
- $f_{Y|x}(y) = f_Y(y); f_{X|y} = f_X(x)$
- $P(X \subset A, y \subset B) = P(X \subset A)P(Y \subset B)$

Covariance & Correlation Covariance: measure dependence between random varibales

$Cov(X, Y) = \delta_{XY} = E(XY) - \mu_X\mu_Y \in (-\infty, \infty)$ If independent, $Cov(X, Y) = 0$. $\rho_{XY} = 0$ is necessary for independence, but not sufficient.

Correlation:

Pearson correlation: normalized covariance to test linear relationship between X and Y, unlikely for broad distribution. $\rho_{XY} = \sigma_{XY}/\sigma_X\sigma_Y \in [-1, 1]$

Spearman rank correlation: test monotonic relationship between X and Y. Calculate ranks (1 to n), $r_X(i)$ and $r_Y(i)$, $Spearman(X, Y) = Pearson(r_X, r_Y)$

Linear functions of random variables

$Y = c_1X_1 + c_2X_2 + ... + C_pX_p$

$E(Y) = c_1E(X_1) + ... + c_pE(X_p)$

$V(Y) = c_1^2V(X_1) + c_p^2V(X_p) + 2\sum_{i<j}\sum c_ic_jcov(X_iX_j)$

If $cov(x_i, x_j) = 0 \rightarrow V(Y) = c_1^2V(X_1) + ... + c_p^2V(X_p)$

Average of X: $\overline{X} = (X_1+X_2+..X_p)/p$, then $E(\overline{X}) = \mu; V(\overline{X}) = \delta^2/p$

Appendix

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967