

Discrete Random Variables

A *dichotomous random variable* takes only the values 0 and 1. Let X be such a random variable, with $\Pr(X=1) = p$ and $\Pr(X=0) = 1-p$. Then $E[X] = p$, and $\text{Var}[X] = p(1-p)$.

Consider a sequence of n independent experiments, each of which has probability p of “being a success.” Let $X_k = 1$ if the k -th experiment is a success, and 0 otherwise. Then the total number of successes in n trials is $X = X_1 + \dots + X_n$; X is a *binomial random variable*, and

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

$E[X] = np$, and $\text{Var}[X] = np(1-p)$. (These results follow from the properties of the expected value and variance of sums of independent random variables.)

Next, consider a sequence of independent experiments, and let Y be the number of trials up to (and including) the first success. Y is a *geometric random variable*, and

$$\Pr(Y = k) = (1-p)^{k-1} p.$$

$E[Y] = 1/p$, and $\text{Var}[Y] = (1-p)/p^2$. (These results follow from the evaluation of infinite sums.)

A *hypergeometric random variable* Z results from drawing a sample of size n from a population of size N containing g “good” members, and then counting the number of “good” members in the sample:

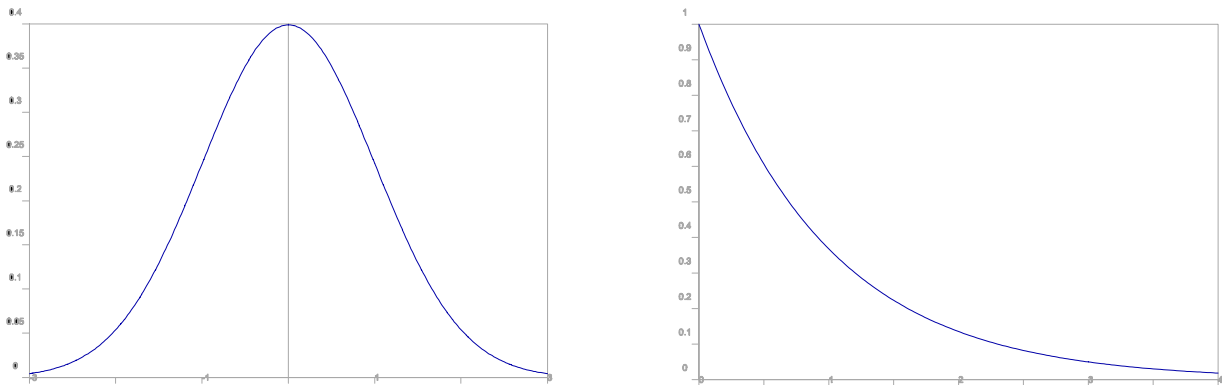
$$\Pr(Z = z) = \frac{\binom{g}{z} \cdot \binom{N-g}{n-z}}{\binom{N}{n}}.$$

(This formula was used to compute the relevant probabilities in the “Bag R vs. Bag B” example.)

Continuous Random Variables

A *continuous random variable* is a random variable which can take any value in some interval. A continuous random variable is characterized by its *probability density function*, a graph which has a total area of 1 beneath it: The probability of the random variable taking values in any interval is simply the area under the curve over that interval.

The **normal distribution**: This most-familiar of continuous probability distributions has the classic “bell” shape (see the left-hand graph below). The peak occurs at the mean of the distribution, i.e., at the expected value of the normally-distributed random variable with this distribution, and the standard deviation (the square root of the variance) indicates the spread of the bell, with roughly 68% of the area within 1 standard deviation of the peak.



The normal distribution arises so frequently in applications due to an amazing fact: If you take a bunch of independent random variables (with comparable variances) and average them, the result will be roughly normally distributed, *no matter what the distributions of the separate variables might be*. (This is known as the “Central Limit Theorem”.) Many interesting quantities (ranging from IQ scores, to demand for a retail product, to lengths of shoelaces) are actually a composite of many separate random variables, and hence are roughly normally distributed.

If X is normal, and $Y = aX + b$, then Y is also normal, with $E[Y] = aE[X] + b$ and $\text{StdDev}[Y] = a \cdot \text{StdDev}[X]$. If X and Y are normal (independent or not), then $X + Y$ and $X - Y = X + (-Y)$ are also normal (intuition: the sum of two bunches is a bunch). Any normally-distributed random variable can be transformed into a “standard” normal random variable (with mean 0 and standard deviation 1) by subtracting off its mean and dividing by its standard deviation. Hence, a single tabulation of the cumulative distribution for a standard normal random variable (attached) can be used to do probabilistic calculations for *any* normally-distributed random variable.

The **exponential distribution**: Consider the time between successive incoming calls at a switchboard, or between successive patrons entering a store. These “interarrival” times are typically exponentially distributed. If the mean interarrival time is $1/\lambda$ (so λ is the mean arrival rate per unit time), then the variance will be $1/\lambda^2$ (and the standard deviation will be $1/\lambda$). The right-hand graph above displays the graph of the exponential density function when $\lambda = 1$. Generally, if X is exponentially distributed, then $\Pr(s < X < t) = e^{-\lambda s} - e^{-\lambda t}$ (where $e \approx 2.71828$).

The exponential distribution fits the examples cited above because it is the only distribution with the “*lack-of-memory*” property: If X is exponentially distributed, then $\Pr(X \leq s+t \mid X > s) = \Pr(X \leq t)$. (After waiting a minute without a call, the probability of a call arriving in the next two minutes is the same as was the probability (a minute ago) of getting a call in the following two minutes. As you continue to wait, the chance of something happening “soon” neither increases nor decreases.) Note that, among discrete distributions, the geometric distribution is the only one with the lack-of-memory property; indeed, the exponential and geometric distributions are analogues of one another.

Let the time between successive arrivals into some system be exponentially distributed, and let N be the number of arrivals in a fixed interval of time of length t . Then N (a discrete random variable) has the ***Poisson distribution***, and

$$\Pr(N = k) = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!} .$$

$E[N] = \lambda t$, and $\text{Var}[N] = \lambda t$ as well. The exponential and Poisson distributions arise frequently in the study of queuing, and of process quality. An interesting (and sometimes useful) fact is that the minimum of two independent, identically-distributed exponential random variables is a new random variable, also exponentially distributed and with a mean precisely half as large as the original mean(s).

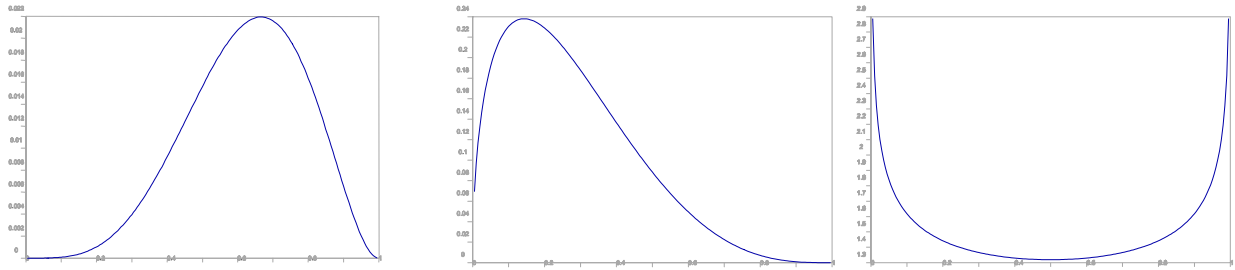
Connections: The geometric distribution deals with the time between successes in a series of independent trials, and the binomial distribution deals with the number of successes in a fixed number of trials. The exponential distribution deals with the time between occurrences of successive events, and the Poisson distribution deals with the number of occurrences in a fixed period of time. Obviously, there's a relationship here. Rule of thumb: If $n > 20$ and $p < 0.05$, then a binomial random variable with parameters (n, p) has a probability distribution very similar to that of a Poisson random variable with parameters $\lambda = np$ and $t = 1$. (Think of dividing one interval of time into n subintervals, and having a probability p of an arrival in each subinterval. That's very much like having a rate of np arrivals (on average) per unit time.)

The *uniform distribution*: A random variable U is uniformly distributed on an interval $[a,b]$ if its density function is flat over that interval. (The uniform distribution is what one typically has in mind when one thinks of “picking a number at random” over some range.) The expected value of U is $(a+b)/2$, and $\text{Var}[U] = (b-a)^2/12$. If n independent, identically distributed random variables are all uniformly distributed on $[0,1]$ (i.e., if we pick n numbers at random between 0 and 1), then the expected values of the largest, second-largest,..., smallest are $n/(n+1)$, $(n-1)/(n+1)$,..., $1/(n+1)$ (i.e., the expected values divide the interval into $n+1$ equally-large subintervals). *Given* that n individuals enter a system (with exponential interarrival times) in a fixed interval of time, the n actual arrival times will look as if they were drawn uniformly over that time interval.

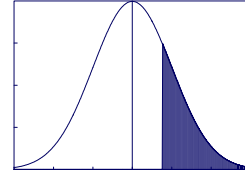
The *beta distribution*: A random variable X has the beta distribution (with parameters $\alpha > 0$ and $\beta > 0$) on the interval $[0, 1]$ if its density has the form $k \cdot x^{\alpha-1}(1-x)^{\beta-1}$ (where k is a scale factor which makes the area under the curve 1).

$$E[X] = \alpha/(\alpha+\beta), \text{ and } \text{Var}[X] = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)] .$$

The main use of the beta distribution is to “fit” it to observed data when building a model of a real-world phenomenon: The beta distribution can take a wide variety of shapes, as seen in the graphs below.



The lefthand graph arises when $\alpha = 5$ and $\beta = 3$, the center when $\alpha = 1.5$ and $\beta = 3$, and the right when $\alpha = 0.5$ and $\beta = 0.5$. Generally, if $\alpha > 2$, the density has a slope of 0 at $x = 0$, if $2 > \alpha > 1$, the density is near-vertical near $x = 0$, and if $1 > \alpha > 0$, the density rises as x approaches 0. (Analogous properties involving β hold when x is near 1.) When $\alpha = \beta = 1$, the beta distribution is uniform on $[0, 1]$.



Right-Tail Probabilities of the Normal Distribution

		+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09	+0.10
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641	0.4602
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247	0.4207
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859	0.3821
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483	0.3446
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121	0.3085
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776	0.2743
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451	0.2420
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148	0.2119
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867	0.1841
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611	0.1587
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379	0.1357
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170	0.1151
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985	0.0968
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823	0.0808
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681	0.0668
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559	0.0548
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455	0.0446
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367	0.0359
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294	0.0287
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233	0.0228
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183	0.0179
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143	0.0139
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110	0.0107
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084	0.0082
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064	0.0062
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048	0.0047
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036	0.0035
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014	0.0013

Simulating Random Variables

Sometimes one wishes to simulate a random process (so as to compare alternative policies on a computer before trying them out in real life). Simulations require the generation of random variables fitting particular distributions. Most computer languages and spreadsheets can provide (pseudo-) random numbers, uniformly distributed between 0 and 1. Fortunately, random variables with *any* probability distribution can be generated from these uniform random variables:

Consider any random variable X . Simply take a uniformly-distributed observation u , solve the equation $\Pr(X \leq x) = u$ for x , and let the resulting x be your observation of X .

For example, in Excel the function `RAND()` returns a random value uniformly distributed between 0 and 1. In addition, the function `NORMINV(,,,.)` returns the inverse of the cumulative normal probability distribution. Therefore, to generate a random observation from a normal distribution with a particular mean and standard deviation, one can simply enter the formula `=NORMINV(RAND(),mean,stddev)` in a cell of the spreadsheet.