PROBABILITY AND RANDOM VARIABLES.

PROBABILITY AND DISTRIBUTION FUNCTIONS.

For discrete random variables, the *probability function* or *probability distribution* $f_X(x)$ such that

$$f_X(x) := P(X = x)$$

is the probability that the random variable X takes on the value x. For continuous random variables, $f_X(x)$ is called the *probability density.* $f_X(x)$ is a non-negative function whose integral over the sample space of the random variable is one.

The *distribution function* $F_X(x)$ such that

$$F_X(x) := P(X \leq x)$$

is the probability that the random variable X takes on values less than or equal to x.

PROBABILITY DISTRIBUTIONS (DISCRETE).

Binomial distribution.

$$f_Z(z) = \binom{n}{z} p^z (1-p)^{n-z}, \quad z \in \mathbb{N}.$$

The distribution of the number of successes z in n independent trials with a fixed probability p of success in each trial, denoted $B(n, p)$.

Poisson distribution.

$$f_Z(z) = \frac{e^{-\lambda}\lambda^z}{z!}, \quad z \in \mathbb{N}.$$

The limit of the binomial distribution under the constraint $np = \lambda$ when $n \to \infty$, so that the mean number of events over a certain time interval is λ. Denoted Poisson(λ).

Geometric distribution.

$$f_Z(z) = p(1-p)^{z-1}, \quad z \in \mathbb{N}.$$

The distribution of the number of independent trials z up to and including the first success with a fixed probability p of success in each trial, denoted Geometric(p).

Bernoulli distribution.

$$f_Z(z) = \begin{cases} p & z = 1, \\ 1-p & z = 0. \end{cases}$$

The distribution of the success of an event. Denoted Bernoulli(p).

PROBABILITY DENSITIES (CONTINUOUS).

Exponential density.

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

The distribution of the time x between consecutive independent events with mean number of events over a certain time interval λ. Denoted Exponential(λ).

Normal density.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

The limiting distribution of a sum of independent random variables with mean μ and variance σ^2, denoted $N(\mu, \sigma^2)$.

Let U be a random variable uniformly distributed on [0, 1), and define $Y = -(1/\lambda)\log U$. Show that when $\lambda > 0$, Y is Exponential(λ).

Manipulating the distribution functions is easier than manipulating the densities, for we have the direct chain of equivalences

$$\begin{aligned} F_Y(y) &= P(Y \geq y) = P(-(1/\lambda)\log U \geq y) \\ &= P(U \geq e^{-\lambda y}) = F_U(e^{-\lambda y}) \\ &= 1 - e^{-\lambda y}, \end{aligned}$$

and taking a derivative with respect to y yields the desired result. Note that there is no such direct relationship for the densities, *i.e.*,

$$f_Y(y) \neq f_U(e^{-\lambda y}).$$

Pick a pattern of heads and tails of length $L < \infty$, and then flip a fair coin repeatedly until the pattern appears. Let N denote the number of flips before the first occurrence of such a pattern, and show that $\langle N \rangle < \infty$.

The distribution of N is difficult to obtain and dependent on the exact sequence chosen because the trials are correlated for $L > 1$. Instead, we construct a finite upper bound to $\langle N \rangle$. Instead of checking every L-trial sequence for a match, we only check the sequences $1 \cdots L$, $L{+}1 \cdots 2L$, $2L{+}1 \cdots 3L$, and so on, so that each trial remains independent even when $L > 1$. Let N' denote the analogue to N in this construction. Our construction is subsumed in the original construction, which checks all the sequences of our construction and more, so we must have $\langle N \rangle \leq \langle N' \rangle$. N'/L is then distributed as Geometric(0.5), so that $\langle N' \rangle = 2^L L < \infty$.

SEVERAL RANDOM VARIABLES.

The above results are readily generalized to larger numbers of random variables, leading to random vectors $\mathbf{x} := (x, y, z, \cdots)$ and multivariate densities $f(\mathbf{x}) := f(x, y, z, \cdots)$. Several additional definitions apply.

Let X and Y be two independent Poisson random variables with respective parameters λ and μ, and define $Z = X + Y$. Show that Z is Poisson($\lambda + \mu$).

We have

$$\begin{aligned} P(Z = z) &= \sum_{x=0}^{z} P(X = x)P(Y = z - x) \\ &= \sum_{x=0}^{z} \frac{\lambda^x e^{-\lambda}}{x!} \frac{\mu^{z-x} e^{-\mu}}{(z-x)!} \\ &= \frac{e^{-(\lambda+\mu)}}{z!} \sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \lambda^x \mu^{z-x} \\ &= \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^z}{z!}, \end{aligned}$$

where the last equality follows from the binomial theorem. Thus Z is indeed Poisson-distributed with parameter $\lambda + \mu$, as was to be shown.

Given invertible distribution functions $F_X(x)$ and $F_Y(y)$ such that $F_X(x) \le F_Y(x)$ for all x, show that there exist random variables X and Y with respective distribution functions F_X and F_Y such that $P(Y \le X) = 1$.

Lemma. $Z = F_Y^{-1}(F_X(X))$ has distribution function $F_Y(z)$. This follows simply from considering the domain and range of each relevant function, with

$$X \xrightarrow{F_X} [0,1] \xrightarrow{F_Y^{-1}} Y,$$

so that Z is distributed according to Y and also parametrized by X.

Hence choosing $X = X$, $Y = F_Y^{-1}(F_X(X))$, we have

$$\begin{aligned} P(Y \le X) &= P(F_Y^{-1}(F_X(X)) \le X) \\ &= P(F_X(X) \le F_Y(X)) = 1, \end{aligned}$$

as desired. The second equality follows because F_Y increases monotonically and hence respects the inequality.

INDEPENDENCE.

Two random variables X and Y are *independent* iff

$$P(X \le x, Y \le y) = P(X \le x) P(Y \le y)$$

for all x and y. Equivalently,

$$F_{X,Y}(x,y) = F_X(x) F_Y(y),$$

and, for f jointly continuous,

$$f_{X,Y}(x,y) = f_X(x) f_Y(y).$$

Let X and Y be random variables with joint density $f_{X,Y}(x,y)$ and sum $Z = X + Y$. Find the probability density $f_Z(z)$.

By hypothesis, we have

$$\begin{aligned} P(Z \le z) &= \int_{x,y\,:\,x+y\le z} \mathrm{d}x\,\mathrm{d}y\, f(x,y) \\ &= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{z-x} \mathrm{d}y\, f(x,y), \end{aligned}$$

so that differentiation with respect to z yields

$$f_Z(z) = \int_{-\infty}^{\infty} \mathrm{d}x\, f_{X,Y}(x, z-x).$$

If f is jointly continuous and x and y are independent, then in addition

$$f_Z(z) = \int_{-\infty}^{\infty} \mathrm{d}x\, f_X(x) f_Y(z-x),$$

so that $f_Z = f_X * f_Y$ is the convolution of f_X and f_Y.

Given the independent exponential random variables X and Y with respective parameters λ and μ, show that $Z = \min(X,Y)$ and the event $E = \{X < Y\}$ are similarly independent. ($P(Z \ge z, E) = P(Z \ge z)P(E)$ iff Z and E are independent, by analogy to the definition for two random variables.)

We have

$$\begin{aligned} P(Z \ge z, E) &= P(X \ge z, E) \\ &= \int_z^{\infty} \mathrm{d}x \int_x^{\infty} \mathrm{d}y\, \lambda\mu e^{-\lambda x - \mu y} \\ &= \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)z}, \end{aligned}$$

and also

$$\begin{aligned} P(Z \ge z)P(E) &= P(X \ge z)P(Y \ge z)P(E) \\ &= \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)z}, \end{aligned}$$

as desired. Note that

$$P(Z \ge z, E) = P(X \ge z, E) \ne P(X \ge z).$$

Let $(X_r;\, r \ge 1)$ be independent random variables with $X_r = \text{Exponential}(\lambda_r)$ and $0 < \sum_{r=1}^{\infty} \lambda_r < \infty$, and let $Y = \inf_{r \ge 1} X_r$. Show that there is, almost surely, a unique N such that $X_N = Y$, and that

$$P(N = n) = \frac{\lambda_n}{\sum_{r=1}^{\infty} \lambda_r}.$$

Equivalently, we must show that

$$1 = P(X_r \ge X_1) + P(X_r \ge X_2) + \cdots$$

for all r. We have

$$\begin{aligned} P(X_r \ge X_1) &= \int_0^{\infty} \mathrm{d}x_1\, \lambda_1 e^{-\lambda_1 x_1} \int_{x_1}^{\infty} \mathrm{d}x_2\, \lambda_2 e^{-\lambda_2 x_2} \cdots \\ &= \int_0^{\infty} \mathrm{d}x_1\, \lambda_1 e^{-\lambda_1 x_1} e^{-\lambda_2 x_1} \cdots = \frac{\lambda_1}{\sum_{r=1}^{\infty} \lambda_r}, \end{aligned}$$

and similarly for the other terms.

MARGINALS.

Given the multivariate density $f_{X,Y}(x,y)$ of the random variables X and Y respectively, the *marginal distribution* of X is

$$f_X(x) = \int_Y \mathrm{d}y\, f_{X,Y}(x,y),$$

where the integral spans the sample space of Y given that $X = x$. In other words, $f_X(x)$ is the probability density for X with Y "integrated out". The marginal distribution of Y is defined similarly.

MULTIVARIATE CHANGE OF VARIABLES.

Given the random variables $U = u(X,Y)$ and $V = v(X,Y)$ for random variables X and Y with joint density $f_{X,Y}(x,y)$, and given invertible, once-differentiable functions u and v, we have from multivariable calculus the standard change-of-variables theorem

$$f_{U,V}(u,v) = f_{X,Y}(x(u,v), y(u,v))|J(u,v)|,$$

where $|J(u,v)|$ is the determinant of the Jacobian $J(u,v)$,

$$J(u,v) \equiv \frac{\partial(x,y)}{\partial(u,v)} := \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}.$$

Given that X and Y are independent Gaussian random variables $N(0,1)$, show that $Z = X/Y$ and Z^{-1} both have the Cauchy density

$$f_Z(z) = \frac{1}{\pi}\frac{1}{1+z^2}.$$

Consider the change of variables $Z = X/Y$, $W = Y$. The multivariate density of X and Y is given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2},$$

and applying the above theorem yields

$$f_{Z,W}(z,w) = f_{X,Y}(zw,w)|w| = \frac{|w|}{2\pi}e^{-w^2(z^2+1)/2}.$$

The marginal distribution $f_Z(z)$ is obtained by integrating out w, giving the desired result. In addition, because X and Y are identical, we may switch X and Y without loss of generality, yielding also the same distribution for $Z^{-1} = Y/X$. Z and Z^{-1} therefore share the same probability density.

EXPECTED VALUE.

Let X be a continuous random variable with probability density $f_X(x)$. The *expected value* or *mean* $\langle X\rangle$ is defined as

$$\langle X\rangle = \int_X \mathrm{d}x\, x f_X(x)$$

given that

$$\int_X \mathrm{d}x\, |x| f_X(x) \quad \text{is finite.}$$

Show that a Cauchy-distributed random variable has an ill-defined mean.

By definition, we have that

$$\int_{-\infty}^{\infty} \mathrm{d}x\, \frac{|x|}{\pi}\frac{1}{1+x^2} = \frac{2}{\pi}\int_0^{\infty}\mathrm{d}x\,\frac{x}{1+x^2} = \infty.$$

Such a random variable has an ill-defined mean despite having a line of symmetry about $x = 0$, the intuitive guess for the mean.

Tail theorem. A non-negative random variable X with distribution function $F_X(x)$ has mean

$$\langle X\rangle = \int_0^{\infty}\mathrm{d}x\, P(X > x) = \int_0^{\infty}\mathrm{d}x\,(1 - F_X(x)).$$

Proof. Begin by defining the *indicator random variable*

$$I(x) = \begin{cases} 1 & X > x, \\ 0 & X \le x, \end{cases}$$

so that $\langle I(x)\rangle = P(X > x) = 1 - F_X(x)$. We further have

$$\int_0^{\infty}\mathrm{d}x\,\langle I(x)\rangle = \left\langle \int_0^{\infty}\mathrm{d}x\, I(x)\right\rangle = \left\langle \int_0^{X}\mathrm{d}x\right\rangle = \langle X\rangle,$$

as desired. The tail theorem relates the expected value $\langle X\rangle$ with the distribution function $F_X(x)$.

Theorem. Let X be a discrete random variable with probability distribution $f_X(x)$, and let $Y = g(X)$. Then

$$\langle Y\rangle = \sum_x g(x) f_X(x),$$

given, in analogy to the definition of expected value, that the sum is well-defined when $g(x)$ is replaced by $|g(x)|$.

Proof.

$$\begin{aligned}\langle Y\rangle &= \sum_y yP(Y = y) \\ &= \sum_y \sum_{x:\ g(x)=y} g(x)P(X = x) = \sum_x g(x) f_X(x).\end{aligned}$$

In the second equality, we note that $P(Y = y)$ is exactly the sum of probabilities $P(X = x)$ such that $g(x) = y$. In the third equality, we note that the sum over y and over $x\colon g(x) = y$ is exactly a sum over all x. Analogous results hold for continuous and multivariate random variables.

The expected value operator is linear, satisfying

$$\langle aX + bY\rangle = a\langle X\rangle + b\langle Y\rangle.$$

MOMENTS.

We define the k^{th} *moment* of the random variable X as $\mu_k = \langle X^k\rangle$ and the k^{th} *central moment* of X as $\sigma_k = \langle (X - \langle X\rangle)^k\rangle$. In particular, the first moment—the mean—is commonly denoted μ, and the second central moment—the variance—is commonly denoted σ^2.

Show that the variance of the sum of independent random variables is the sum of the variance of each random variable.

Let X and Y be independent random variables and Z be their sum. We first note the identity

$$\sigma_Z^2 = \langle (Z - \langle Z\rangle)^2\rangle = \langle Z^2\rangle + \langle Z\rangle^2 - 2\langle Z\rangle^2 = \langle Z^2\rangle - \langle Z\rangle^2,$$

from which we have

$$\begin{aligned}\sigma_Z^2 &= \langle Z^2\rangle - \langle Z\rangle^2 \\ &= \langle X^2 + 2XY + Y^2\rangle - (\langle X\rangle + \langle Y\rangle)^2 \\ &= \langle X^2\rangle - \langle X\rangle^2 + \langle Y^2\rangle - \langle Y\rangle^2 = \sigma_X^2 + \sigma_Y^2,\end{aligned}$$

as desired. We made use of the independence of the two variables to set $2\langle XY\rangle - 2\langle X\rangle\langle Y\rangle = 0$.

In considering pairs of random variables, labeled X and Y, we further define the *covariance*

$$\text{cov}(X,Y) = \langle (X - \langle X\rangle)(Y - \langle Y\rangle)\rangle,$$

as well as the *correlation coefficient* or *correlation*

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}.$$

Two random variables with zero correlation (hence also zero covariance) are *uncorrelated.*

Show that independent random variables are uncorrelated.

Let X and Y be two independent random variables. Then

$$\text{cov}(X,Y) = \langle XY\rangle - 2\langle X\rangle\langle Y\rangle + \langle X\rangle\langle Y\rangle = 0.$$

Uncorrelated variables, however, are not necessarily independent.

CONDITIONING.

For X and Y jointly discrete random variables, the *conditional probability distribution* of X given Y is defined as

$$f_{X|Y}(x \mid y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Similarly, the *conditional expectation* of X given Y is

$$\langle X \mid Y\rangle = \left(\sum_{x,y|Y} x f_{X,Y}(x,y)\right) \Big/ \left(\sum_{x,y|Y} f_{X,Y}(x,y)\right).$$

Conditional distributions, expectations, moments, and correlations all satisfy the same properties as their regular counterparts. These definitions are also readily generalized to continuous random variables.

We further define two random variables X and Y to be *conditionally independent* given Z if

$$F_{X,Y|Z}(x,y \mid z) = F_{X|Z}(x \mid z) F_{Y|Z}(y \mid z).$$

Let X and Y be independent continuous random variables. Find $P(X < Y)$.

$$\begin{aligned} P(X<Y) &= \int_Y \mathrm{d}y\, P(X<Y \mid Y=y) f_Y(y) \\ &= \int_Y \mathrm{d}y\, F_X(y) f_Y(y). \end{aligned}$$

Identity. For X and Y continuous random variables,

$$\begin{aligned} \langle X\rangle &= \int_X \mathrm{d}x\, x f_X(x) = \int_X \mathrm{d}x \int_Y \mathrm{d}y\, x f_{X,Y}(x,y) \\ &= \int_X \mathrm{d}x \int_Y \mathrm{d}y\, x f_{X|Y}(x \mid Y=y) f_Y(y) \\ &= \int_Y \mathrm{d}y\, \langle X \mid Y=y\rangle f_Y(y) = \langle\langle X \mid Y\rangle_X\rangle_Y. \end{aligned}$$

The identity holds true also for discrete random variables, but the integrals must be replaced with corresponding sums. The subscripts on the angle brackets make clear which variables the expected values are taken with respect to.

Let X and Y be independent continuous random variables. Find $P(X < Y)$.

We will use our newly derived identity. Define the indicator variable $I_{X<Y}$, which is 1 when $X < Y$ and 0 otherwise. We note that $\langle I\rangle = P(X < Y)$ and that $\langle I \mid Y = y\rangle = P(X < y) = F_X(y)$, so

$$P(X<Y) = \langle I\rangle = \langle\langle I \mid Y\rangle_I\rangle_Y = \langle F_X(y)\rangle,$$

as has been found.

Pull-through property. Given random variables X and Y,

$$\langle X g(Y) \mid Y=y\rangle = \langle X g(y) \mid Y=y\rangle = g(y)\langle X \mid Y=y\rangle.$$

Tower property. Given jointly distributed random variables X, Y, and Z,

$$\langle\langle X \mid Y, Z\rangle \mid Z\rangle = \langle X \mid Z\rangle,$$

which follows from the identity $\langle\langle X \mid Y\rangle\rangle = \langle X\rangle$ when all the expectation values are conditioned on Z.

GENERATING FUNCTIONS.

The *moment-generating function (mgf)* of a random variable X is defined as

$$M_X(t) = \langle e^{tX}\rangle$$

for all real t such that $M_X(t)$ exists. Within a disk of convergence centered about the origin, we have

$$M_X(t) = \langle e^{tX}\rangle = \sum_{n=0}^{\infty} \frac{t^n}{n!}\langle X^n\rangle,$$

so that the moments of X can be obtained as

$$\langle X^n\rangle = \frac{\mathrm{d}^n M_X}{\mathrm{d}t^n}(0).$$

Moment-generating functions are unique: to every probability distribution is associated a unique mgf, and *vice versa.* This correspondence can be constructed directly via Laplace transformation and inversion, with

$$M_X(t) = \langle e^{tX}\rangle = \int_0^\infty \mathrm{d}x\, e^{tx} f_X(x) = \mathcal{L}[f_X(x)](-t),$$

and hence that

$$f_X(x) = \mathcal{L}^{-1}[M_X(-t)](x).$$

If X is defined over the real line and not on the positive half-line, then we use the bilateral Laplace transform with lower integral bound $-\infty$ instead of 0. Additionally, moment-generating functions have the continuity property: to a limit of probability distributions $p_n(x) \to p(x)$ is correspondingly associated the mgfs $M_{X_n}(t) \to M_X(t)$. The proof is difficult and omitted.

Show that

$$\langle X^{-1}\rangle = \int_0^\infty \mathrm{d}t\, M_X(-t).$$

We have

$$\langle X^{-1}\rangle = \left\langle \int_0^\infty \mathrm{d}t\, e^{-tX} \right\rangle = \int_0^\infty \mathrm{d}t\, M_X(-t).$$

For jointly distributed random variables X and Y, we can define the analogous joint mgf

$$M_{X,Y}(s,t) = \langle e^{sX+tY}\rangle,$$

from which singular and joint moments can be obtained by respective partial derivatives. If X and Y are independent, then we further have

$$M_{X,Y}(s,t) = \langle e^{sX+tY}\rangle = \langle e^{sX}\rangle\langle e^{tY}\rangle = M_X(s)M_Y(t).$$

For discrete, integer-valued random variables, we also define the *probability-generating function (pgf)*

$$G_X(s) = \langle s^X\rangle = \sum_n s^n f_X(n)$$

for such a random variable X, satisfying $G_X(e^t) = M_X(t)$. The probability-generating function is a variant of the moment-generating function with scope limited to such variables.

Let the random variable N be Poisson(λ), X be Binomial(N,p), and $Y = N - X$. Show that X and Y are independent and find the distribution of each.

We have

$$\begin{aligned}
\langle s^X\rangle &= \sum_{n=0}^\infty \sum_{x=0}^n s^x P(N=n, X=x) \\
&= \sum_{n=0}^\infty \frac{e^{-\lambda}\lambda^n}{n!} \sum_{x=0}^n \binom{n}{x} p^x(1-p)^{n-x} s^x \\
&= e^{-\lambda}e^{\lambda(ps+1-p)} = e^{\lambda p(s-1)}, \\
\langle t^Y\rangle &= \sum_{n=0}^\infty \sum_{x=0}^n t^{n-x} P(N=n, X=x) \\
&= \sum_{n=0}^\infty \frac{e^{-\lambda}\lambda^n}{n!} \sum_{x=0}^n \binom{n}{x} p^x(1-p)^{n-x} t^{n-x} \\
&= e^{-\lambda}e^{\lambda(p+(1-p)t)} = e^{\lambda(1-p)(t-1)}, \\
\langle s^X t^Y\rangle &= \sum_{n=0}^\infty \sum_{x=0}^n s^x t^{n-x} P(N=n, X=x) \\
&= \sum_{n=0}^\infty \frac{e^{-\lambda}\lambda^n}{n!} \sum_{x=0}^n \binom{n}{x} p^x(1-p)^{n-x} s^x t^{n-x} \\
&= e^{-\lambda}e^{\lambda(ps+(1-p)t)} = \langle s^X\rangle\langle t^Y\rangle,
\end{aligned}$$

so indeed X and Y are independent, X being Poisson(λp) and Y being Poisson($\lambda(1-p)$).

Finally, we define the *characteristic function (cf)*

$$\phi_X(t) = \langle e^{itX}\rangle,$$

which duplicates the properties of the moment-generating function and always exists, even when the moment-generating function may not, because

$$|\langle e^{itX}\rangle| \le \langle |e^{itX}|\rangle = 1,$$

which itself follows from the properties of integration.

Find the characteristic function of the Cauchy distribution. Show that the corresponding moment-generating function does not exist apart from the origin.

We have shown earlier that the mean of the Cauchy distribution is ill-defined, and so too must be its moment-generating function; this can also be shown by direct evaluation. The characteristic function of this distribution is given by

$$\langle e^{itX}\rangle = \int_{-\infty}^\infty \mathrm{d}x\, \frac{1}{\pi}\frac{e^{itx}}{1+x^2} = e^{-|t|}$$

by Cauchy's residue theorem.

Uniqueness follows from Fourier transformation and inversion, with

$$\phi_X(t) = \langle e^{itX}\rangle = \int_{-\infty}^\infty \mathrm{d}x\, e^{itx} f_X(x) = \mathcal{F}[f_X(x)](t)$$

and also

$$f_X(x) = \mathcal{F}^{-1}[\phi_X(t)](x).$$

Lévy's continuity theorem guarantees the continuity property; the proof is again omitted.

The cf and the mgf are related by analytic continuation in the complex plane, with $\phi_X(t) = M_X(it)$ subject to suitable constraints. Given $M_X(t)$ on the interval $t \in (-a,a)$, then $\phi_X(t)$ is as stated. Given $\phi_X(z)$ differentiable in the disk $|z| < a$, then $M_X(t)$ is as stated in the interval $t \in (-a,a)$.

INTRODUCTION TO STOCHASTIC PROCESSES.

SIMPLE RESULTS.

Markov inequality. Let X be a non-negative random variable. Then, for any $a > 0$,

$$P(X \geq a) \leq \frac{\langle X \rangle}{a}.$$

Proof. Define the indicator

$$I(a) = \begin{cases} 1, & X \geq a, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly $aI(a) \leq X$, so that $aP(X > a) = a\langle I(a) \rangle \leq \langle X \rangle$.

Chebyshev inequality. Let X be a random variable. Then, for any $a > 0$,

$$P(|X| \geq a) \leq \frac{\langle X^2 \rangle}{a^2}.$$

Proof. Apply the Markov inequality with $a \to a^2$ to the random variable X^2, and note that $X^2 \geq a^2$ implies $|X| \geq a$.

Borel-Cantelli lemma. Let $(A_n; n \geq 1)$ be a collection of events, and let A be the event $\{A_n \text{ i.o.}\}$ that infinitely many of the A_n occur. If $\sum_{n=0}^{\infty} P(A_n) < \infty$, then $P(A) = 0$.

Proof.

$$A = \bigcap_{m=1}^{\infty} \bigcup_{r=m}^{\infty} A_r \subseteq \bigcup_{r=m}^{\infty} A_r.$$

Hence

$$P(A) \leq P\left(\bigcup_{r=m}^{\infty} A_r\right) \leq \sum_{r=m}^{\infty} P(A_r)$$

for all m, and the rightmost expression approaches 0 as $m \to \infty$.

Second Borel-Cantelli lemma. Let $(A_n; n \geq 1)$ be a collection of independent events with $\sum_{n=1}^{\infty} P(A_n) = \infty$. Then $P(A_n \text{ i.o.}) = 1$.

The proof is omitted.

ISSUES OF CONVERGENCE.

Summation lemma. Define the event $A_n(\epsilon) = \{|X_n - X| > \epsilon\}$ for $\epsilon > 0$ for a stochastic process $(X_n; n \geq 0)$ and a limit X. As $n \to \infty$, $P(X_n \to X) = 1$ if and only if finitely many $A_n(\epsilon)$ occur for any ϵ.

The proof is omitted. Note that, by application of the Borel-Cantelli lemma, $\sum_{n=0}^{\infty} P(A_n(\epsilon)) < \infty$ implies that $P(X_n \to X) = 1$.

Convergence almost surely. $P(X_n \to X) = 1$ as $n \to \infty$ defines the strong *almost sure convergence*, denoted

$$X_n \xrightarrow{\text{a.s.}} X.$$

Convergence in probability. If, for all $\epsilon > 0$,

$$P(A_n(\epsilon)) = P(|X_n - X| > \epsilon) \to 0 \quad \text{as} \quad n \to \infty,$$

then X_n *converges in probability* to X and we write

$$X_n \xrightarrow{\text{P}} X.$$

Convergence in probability is weaker than convergence almost surely, for the former is implied by the latter. The inverse, however, is not true: convergence in probability has no constraints on the "rate" of convergence, and too slow a rate of convergence may lead to convergence in probability but not almost surely.

> Construct a stochastic process that converges in probability but not almost surely.
>
> ---
>
> Let each U_n be distributed uniformly on $[0, 1)$, and let the X_n be indicator variables such that
>
> $$X_n = \begin{cases} 1, & U_n < 1/n \\ 0, & \text{otherwise.} \end{cases}$$
>
> Then $\lim_{n\to\infty} P(X_n = 0) = 1$, so the limit $X = 0$ and
>
> $$P(|X_n - X| > \epsilon) = P(X_n > \epsilon) = \begin{cases} 1/n, & \epsilon < 1 \\ 0, & \text{otherwise.} \end{cases}$$
>
> Therefore
>
> $$X_n \xrightarrow{\text{P}} X.$$
>
> On the other hand, for any $\epsilon \in [0, 1)$ and by the second Borel-Cantelli lemma, we have $P(A_n \text{ i.o.}) = 1$. Then, by the summation lemma, $P(X_n \to X) \neq 1$, so X_n does not converge almost surely to X. The key insight in this construction is that the harmonic series diverges, suggesting that a stochastic process with relevant probabilities being the terms in the harmonic series will satisfy the required criteria.

Convergence in mean square. By Chebyshev's inequality,

$$P(|X_n - X| > \epsilon) \leq \frac{\langle |X_n - X|^2 \rangle}{\epsilon^2},$$

so that a sequence (X_n) converges in probability if $\lim_{n\to\infty} \langle |X_n - X|^2 \rangle = 0$. This latter criterion defines *convergence in mean square*, and is denoted

$$X_n \xrightarrow{\text{m.s.}} X.$$

Convergence in mean square implies convergence in probability, but not *vice versa*.

> Construct a stochastic process that converges in probability but not in mean square.
>
> ---
>
> Consider the previous example given that
>
> $$X_n = \begin{cases} n, & U_n < 1/n \\ 0, & \text{otherwise.} \end{cases}$$

Convergence in mean square neither implies nor is implied by convergence almost surely.

> Construct a stochastic process that converges in mean square but not almost surely.
>
> ---
>
> Consider the counterexample regarding convergence in probability but not almost surely.