4. Expectation

4.1 Expectation of a Random Variable

The **expected value**, **mean** or **first moment** of X is defined to be

$$\mathbb{E}(X) = \int x \; dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) \; dx & \text{if } X \text{ is continuous} \end{cases}$$

assuming that the sum (or integral) is well-defined. We use the following notation to denote the expected value of X:

$$\mathbb{E}(X) = \mathbb{E}X = \int x \; dF(x) = \mu = \mu_X$$

The expectation is a one-number summary of the distribution. Think of $\mathbb{E}(X)$ as the average value you'd obtain if you computed the numeric average $n^{-1} \sum_{i=1}^n X_i$ for a large number of IID draws $X_1, \dots, X_n$. The fact that $\mathbb{E}(X) \approx n^{-1} \sum_{i=1}^n X_i$ is a theorem called the law of large numbers which we will discuss later. We use $\int x \; dF(x)$ as a convenient unifying notation between the discrete case $\sum_x x f(x)$ and the continuous case $\int x f(x) \; dx$ but you should be aware that $\int x \; dF(x)$ has a precise meaning discussed in real analysis courses.

To ensure that $\mathbb{E}(X)$ is well defined, we say that $\mathbb{E}(X)$ exists if $\int_x |x| \; dF_X(x) < \infty$. Otherwise we say that the expectation does not exist. From now on, wheneverwe discuss expectations, we implicitly assume they exist.

Theorem 4.6 (The rule of the lazy statician). Let $Y = r(X)$. Then

$$\mathbb{E}(Y) = \mathbb{E}(r(X)) = \int r(x) \; dF_X(x)$$

As a special case, let A be an event and let $r(x) = I_A(x)$, where $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ otherwise. Then

$$\mathbb{E}(I_A(X)) = \int I_A(x) f_X(x) dx = \int_A f_X(x) dx = \mathbb{P}(X \in A)$$

In other words, probability is a special case of expectation.

Functions of several variables are handled in a similar way. If $Z = r(X, Y)$ then

$$\mathbb{E}(Z) = \mathbb{E}(r(X, Y)) = \int \int r(x, y) \; dF(x, y)$$

The k**-th moment** of X is defined to be $\mathbb{E}(X^k)$, assuming that $\mathbb{E}(|X|^k) < \infty$. We shall rarely make much use of moments beyond $k = 2$.

4.2 Properties of Expectations

Theorem 4.10. If $X_1, \dots, X_n$ are random variables and $a_1, \dots, a_n$ are constants, then

$$\mathbb{E}\left(\sum_i a_i X_i\right) = \sum_i a_i \mathbb{E}(X_i)$$

Theorem 4.12. Let $X_1, \dots, X_n$ be independent random variables. Then,

$$\mathbb{E}\left(\prod_i X_i\right) = \prod_i \mathbb{E}(X_i)$$

Notice that the summation rule does not require independence but the product does.

4.3 Variance and Covariance

Let X be a random variable with mean μ. The **variance** of X -- denoted by σ^2 or σ_X^2 or $\mathbb{V}(X)$ or $\mathbb{V}X$ -- is defined by

$$\sigma^2 = \mathbb{E}(X - \mu)^2 = \int (x - \mu)^2 \, dF(x)$$

assuming this expectation exists. The **standard deviation** is $\text{sd}(X) = \sqrt{\mathbb{V}(X)}$ and is also denoted by σ and σ_X.

Theorem 4.14. Assuming the variance is well defined, it has the following properties:

1. $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
2. If a and b are constants then $\mathbb{V}(aX + b) = a^2 \mathbb{V}(X)$
3. If $X_1, \dots, X_n$ are independent and $a_1, \dots, a_n$ are constants then

$$\mathbb{V}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \mathbb{V}(X_i)$$

If $X_1, \dots, X_n$ are random variables then we define the **sample mean** to be

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the **sample variance** to be

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

Theorem 4.16. Let $X_1, \dots, X_n$ be IID and let $\mu = \mathbb{E}(X_i)$, $\sigma^2 = \mathbb{V}(X_i)$. Then

$$\mathbb{E}\left(\overline{X}_n\right) = \mu, \quad \mathbb{V}\left(\overline{X}_n\right) = \frac{\sigma^2}{n}, \quad \text{and} \quad \mathbb{E}\left(S_n^2\right) = \sigma^2$$

If X and Y are random variables, then the covariance and correlation between X and Y measure how strong the linear relationship between X and Y is.

Let X and Y be random variables with means μ_X and μ_Y and standard deviation σ_X and σ_Y. Define the **covariance** between X and Y by

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

and the **correlation** by

$$\rho = \rho_{X,Y} = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Theorem 4.18. The covariance satisfies:

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

The correlation satisfies:

$$-1 \leq \rho(X, Y) \leq 1$$

If $Y = a + bX$ for some constants a and b then $\rho(X, Y) = 1$ if $b > 0$ and $\rho(X, Y) = -1$ if $b < 0$. If X and Y are independent, then $\text{Cov}(X, Y) = \rho = 0$. The converse is not true in general.

Theorem 4.19.

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\text{Cov}(X, Y) \quad \text{and} \quad \mathbb{V}(X - Y) = \mathbb{V}(X) + \mathbb{V}(Y) - 2\text{Cov}(X, Y)$$

More generally, for random variables $X_1, \ldots, X_n$,

$$\mathbb{V}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \mathbb{V}(X_i) + 2 \sum \sum_{i<j} a_i a_j \text{Cov}(X_i, X_j)$$

4.4 Expectation and Variance of Important Random Variables

Distribution	Mean	Variance
Point mass at p	a	0
Bernoulli(p)	p	$p(1-p)$
Binomial(n, p)	np	$np(1-p)$
Geometric(p)	$1/p$	$(1-p)/p^2$
Poisson(λ)	λ	λ
Uniform(a, b)	$(a+b)/2$	$(b-a)^2/12$
Normal(μ, σ^2)	μ	σ^2
Exponential(β)	β	β^2
Gamma(α, β)	$\alpha\beta$	$\alpha\beta^2$
Beta(α, β)	$\alpha/(\alpha+\beta)$	$\alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$
t_ν	0 (if $\nu > 1$)	$\nu/(\nu-2)$ (if $\nu > 2$)
χ^2_p	p	$2p$
Multinomial(n, p)	np	see below
Multivariate Nornal(μ, Σ)	μ	Σ

The last two entries in the table are multivariate models which involve a random vector X of the form

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$$

The mean of a random vector X is defined by

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_k) \end{pmatrix}$$

The **variance-covariance matrix** Σ is defined to be

$$\Sigma = \begin{pmatrix} \mathbb{V}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_k) \\ \text{Cov}(X_2, X_1) & \mathbb{V}(X_2) & \cdots & \text{Cov}(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_k, X_1) & \text{Cov}(X_k, X_2) & \cdots & \mathbb{V}(X_k) \end{pmatrix}$$

If $X \sim \text{Multinomial}(n, p)$ then

$$\mathbb{E}(X) = np = n(p_1, \dots, p_k) \quad \text{and} \quad \mathbb{V}(X) = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \cdots & np_1p_k \\ -np_2p_1 & np_2(1-p_2) & \cdots & np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \cdots & np_k(1-p_k) \end{pmatrix}$$

To see this:

- Note that the marginal distribution of any one component is $X_i \sim \text{Binomial}(n, p_i)$, so $\mathbb{E}(X_i) = np_i$ and $\mathbb{V}(X_i) = np_i(1-p_i)$.
- Note that, for $i \neq j$, $X_i + X_j \sim \text{Binomial}(n, p_i + p_j)$, so $\mathbb{V}(X_i + X_j) = n(p_i + p_j)(1 - (p_i + p_j))$.
- Using the formula for the covariance of a sum, for $i \neq j$,

$$\mathbb{V}(X_i + X_j) = \mathbb{V}(X_i) + \mathbb{V}(X_j) + 2\text{Cov}(X_i, X_j) = np_i(1-p_i) + np_j(1-p_j) + 2\text{Cov}(X_i, X_j)$$

Equating the last two formulas we get a formula for the covariance, $\text{Cov}(X_i, X_j) = -np_ip_j$.

Finally, here's a lemma that can be useful for finding means and variances of linear combinations of multivariate random vectors.

Lemma 4.20. If a is a vector and X is a random vector with mean μ and variance Σ then

$$\mathbb{E}(a^TX) = a^T\mu \quad \text{and} \quad \mathbb{V}(a^TX) = a^T\Sigma a$$

If A is a matrix then

$$\mathbb{E}(AX) = A\mu \quad \text{and} \quad \mathbb{V}(AX) = A\Sigma A^T$$

4.5 Conditional Expectation

The conditional expectation of X given $Y = y$ is

$$\mathbb{E}(X|Y=y) = \begin{cases} \sum x f_{X|Y}(x|y) & \text{discrete case} \\ \int x f_{X|Y}(x|y)dy & \text{continuous case} \end{cases}$$

If r is a function of x and y then

$$\mathbb{E}(r(X,Y)|Y=y) = \begin{cases} \sum r(x,y) f_{X|Y}(x|y) & \text{discrete case} \\ \int r(x,y) f_{X|Y}(x|y)dy & \text{continuous case} \end{cases}$$

While $\mathbb{E}(X)$ is a number, $\mathbb{E}(X|Y=y)$ is a function of y. Before we observe Y, we don't know the value of $\mathbb{E}(X|Y=y)$ so it is a random variable which we denote $\mathbb{E}(X|Y)$. In other words, $\mathbb{E}(X|Y)$ is the random variable whose value is $\mathbb{E}(X|Y=y)$ when Y is observed as y. Similarly, $\mathbb{E}(r(X,Y)|Y)$ is the random variable whose value is $\mathbb{E}(r(X,Y)|Y=y)$ when Y is observed as y.

Theorem 4.23 (The rule of iterated expectations). For random variables X and Y, assuming the expectations exist, we have that

$$\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}(Y) \quad \text{and} \quad \mathbb{E}[\mathbb{E}(X|Y)] = \mathbb{E}(X)$$

More generally, for any function $r(x,y)$ we have

$$\mathbb{E}[\mathbb{E}(r(X,Y)|X)] = \mathbb{E}(r(X,Y)) \quad \text{and} \quad \mathbb{E}[\mathbb{E}(r(X,Y)|Y)] = \mathbb{E}(r(X,Y))$$

Proof. We will prove the first equation.

$$\begin{aligned} \mathbb{E}[\mathbb{E}(Y|X)] &= \int \mathbb{E}(Y|X=x) f_X(x)dx = \int \int y f(y|x)dy f(x)dx && (1) \\ &= \int \int y f(y|x) f(x) dx dy = \int \int y f(x,y) dx dy = \mathbb{E}(Y) && (2) \end{aligned}$$

The **conditional variance** is defined as

$$\mathbb{V}(Y|X=x) = \int (y - \mu(x))^2 f(y|x)dx$$

where $\mu(x) = \mathbb{E}(Y|X=x)$.

Theorem 4.26. For random variables X and Y,

$$\mathbb{V}(Y) = \mathbb{E}\mathbb{V}(Y|X) + \mathbb{V}\mathbb{E}(Y|X)$$

4.6 Technical Appendix

4.6.1 Expectation as an Integral

The integral of a measurable function $r(x)$ is defined as follows. First suppose that r is simple, meaning that it takes finitely many values $a_1, \ldots, a_k$ over a partition $A_1, \ldots, A_k$. Then $\int r(x) dF(x) = \sum_{i=1}^k a_i \mathbb{P}(r(X) \in A_i)$. The integral of a positive measurable function r is defined by $\int r(x) dF(x) = \lim_i \int r_i(x) dF(x)$, where r_i is a sequence of simple functions such that $r_i(x) \leq r(x)$ and $r_i(x) \to r(x)$ as $i \to \infty$. This does not depend on the particular sequence. The integral of a measurable function r is defined to be $\int r(x) dF(x) = \int r^+(x) dF(x) - \int r^-(x) dF(x)$ assuming both integrals are finite, where $r^+(x) = \max\{r(x), 0\}$ and $r^-(x) = \min\{r(x), 0\}$.

4.6.2 Moment Generating Functions

The **moment generating function (mgf)** or **Laplace transform** of X is defined by

$$\psi_X(t) = \mathbb{E}(e^{tX}) = \int e^{tx} dF(x)$$

where t varies over the real numbers.

In what follows, we assume the mgf is well defined for all t in small neighborhood of 0. A related function is the characteristic function, defined by $\mathbb{E}(e^{itX})$ where $i = \sqrt{-1}$. This function is always defined for all t. The mgf is useful for several reasons. First, it helps us compute the moments of a distribution. Second, it helps us find the distribution of sums of random variables. Third, it is used to prove the central limit theorem.

When the mgf is well defined, it can be shown that we can interchange the operations of differentiation and "taking expectation". This leads to

$$\psi'(0) = \left[\frac{d}{dt}\mathbb{E}e^{tX}\right]_{t=0} = \mathbb{E}\left[\frac{d}{dt}e^{tX}\right]_{t=0} = \mathbb{E}[Xe^{tX}]_{t=0} = \mathbb{E}(X)$$

By taking further derivatives we conclude that $\psi^{(k)}(0) = \mathbb{E}(X^k)$. This gives us a method for computing the moments of a distribution.

Lemma 4.30. Properties of the mgf.

1. If $Y = aX + b$ then $\psi_Y(t) = e^{bt}\psi_X(at)$
2. if $X_1, \ldots, X_n$ are independent and $Y = \sum_i X_i$ then $\psi_Y(t) = \prod_i \psi_i(t)$, where ψ_i is the mgf of X_i.

Theorem 4.32. Let X and Y be random variables. If $\psi_X(t) = \psi_Y(t)$ for all t in an open interval around 0, then $X \overset{d}{=} Y$.

Moment Generating Function for Some Common Distributions

Distribution	mgf
$\text{Bernoulli}(p)$	$pe^t + (1-p)$
$\text{Binomial}(n,p)$	$(pe^t + (1-p))^n$
$\text{Poisson}(\lambda)$	$e^{\lambda(e^t - 1)}$
$\text{Normal}(\mu, \sigma^2)$	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$
$\text{Gamma}(\alpha, \beta)$	$\left(\frac{\beta}{\beta - t}\right)^\alpha$ for $t < \beta$

4.7 Exercises

Exercise 4.7.1. Suppose we play a game where we start with c dollars. On each play of the game you either double your money or half your money, with equal probability. What is your expected fortune after n trials?

Solution. Let the random variables X_i be the fortune after the i-th trial, $X_0 = c$ always taking the value c. Then:

$$\mathbb{E}[X_{i+1} | X_i = x] = 2x \cdot \frac{1}{2} + \frac{x}{2} \cdot \frac{1}{2} = \frac{5}{4} x$$

Taking the expectation on X_i on both sides (i.e. integrating over $F_{X_i}(x)$),

$$\mathbb{E}(\mathbb{E}[X_{i+1} | X_i = x]) = \frac{5}{4}\mathbb{E}(X_i) \Longrightarrow \mathbb{E}(X_{i+1}) = \frac{5}{4}\mathbb{E}(X_i)$$

Therefore, by induction,

$$\mathbb{E}(X_n) = \left(\frac{5}{4}\right)^n c$$

Note that this is **not** a martingale, as in the traditional double-or-nothing formulation -- the expected value goes up at each iteration.

Exercise 4.7.2. Show that $\mathbb{V}(X) = 0$ if and only if there is a constant c such that $\mathbb{P}(X = c) = 1$.

Solution. We have $\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$:

$$\mathbb{V}(X) = \int (x - \mu_X)^2 dF_X(x)$$

Since $(x - \mu_X)^2 \geq 0$, in order for the variance to be 0 we must have the integrand be zero with probability 1, i.e. $\mathbb{P}(X = \mu_X) = 1$.

Exercise 4.7.3. Let $X_1, \dots, X_n \sim \text{Uniform}(0, 1)$ and let $Y_n = \max\{X_1, \dots, X_n\}$. Find $\mathbb{E}(Y_n)$.

Solution. The CDF of Y_n, for $0 \leq y \leq 1$, is:

$$F_{Y_n}(y) = \mathbb{P}(Y_n \leq y) = \prod_{i=1}^n \mathbb{P}(X_i \leq y) = y^n$$

so its PDF is $f_{Y_n}(y) = F'_{Y_n}(y) = ny^{n-1}$ for $0 \leq y \leq 1$.

The expected value of Y_n then is

$$\mathbb{E}(Y_n) = \int_0^1 y f_{Y_n}(y) dy = \int_0^1 ny^n dy = \frac{n}{n+1}$$

Exercise 4.7.4. A particle starts at the origin of the real line and moves along the line in jumps of one unit. For each jump the probability is p that the particle will move one unit to the left and the probability is $1 - p$ that the particle will jump one unit to the right. Let X_n be the position of the particle after n units. Find $\mathbb{E}(X_n)$ and $\mathbb{V}(X_n)$. (This is known as a random walk.)

Solution.

We can define $X_n = \sum_{i=1}^n (1 - 2Y_i)$, where $Y_i \sim \text{Bernoulli}(p)$ and the Y_i's are independent random variables representing the direction of each jump.

We then have:

$$\mathbb{E}(X_n) = \sum_{i=1}^n \mathbb{E}(1 - 2Y_i) = \sum_{i=1}^n (1 - 2p) = n(1 - 2p)$$

and

$$\mathbb{V}(X_n) = \sum_{i=1}^n \mathbb{V}(1 - 2Y_i) \sum_{i=1}^n 4\mathbb{V}(Y_i) = 4np(1 - p)$$

Exercise 4.7.5. A fair coin is tossed until a head is obtained. What is the expected number of tosses that will be required?

Solution. The number of tosses follows a geometric distribution, $X \sim \text{Geom}(p)$, where p is the probability of heads. Let's deduce its expected value, rather than use it as a known fact ($\mathbb{E}(X) = 1/p$). The PDF is

$$f_X(k) = p(1-p)^{k-1}, \quad k > 0$$

The expected value for X is

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^{\infty} kp(1-p)^{k-1} && (3)\\
&= \sum_{k=1}^{\infty} p(1-p)^{k-1} + \sum_{k=2}^{\infty} (k-1)p(1-p)^{k-1} && (4)\\
&= p\left(1 + (1-p) + (1-p)^2 + \ldots\right) + \sum_{k=1}^{\infty} kp(1-p)^{k} && (5)\\
&= p\left(\frac{1}{1-(1-p)}\right) + (1-p)\sum_{k=1}^{\infty} kp(1-p)^{k-1} && (6)\\
&= 1 + (1-p)\mathbb{E}(X) && (7)
\end{aligned}$$

from where we get $\mathbb{E}(X) = 1/p$.

Exercise 4.7.6. Prove Theorem 4.6 for discrete random variables.

Let $Y = r(X)$. Then

$$\mathbb{E}(Y) = \mathbb{E}(r(X)) = \int r(x)\ dF_X(x)$$

Solution. The result is immediate from the definition of expectation:

$$Y(\omega) = r(X(\omega)) = r(x) \quad \forall \omega : X(\omega) = x$$

and so

$$\mathbb{E}(Y) = \int r(x) dF_x(x)$$

Exercise 4.7.7. Let X be a continuous random variable with CDF F. Suppose that $\mathbb{P}(X > 0) = 1$ and that $\mathbb{E}(X)$ exists. Show that $\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > x) dx$.

Hint: Consider integrating by parts. The following fact is helpful: if $\mathbb{E}(X)$ exists then $\lim_{x \to +\infty} x|1 - F(x)| = 0$.

Solution. Let's prove the following, slightly more general, lemma.

Lemma: For every continuous random variable X,

$$\mathbb{E}(X) = \int_0^\infty (1 - F_X(y)) dy - \int_{-\infty}^0 F_X(y) dy$$

Proof:

$$\begin{align}
\mathbb{E}(X) &= \int_{-\infty}^\infty x f_X(x) dx \tag{8} \\
&= \int_{-\infty}^0 \int_x^0 -f_X(x) dy dx + \int_0^\infty \int_0^x f_X(x) dy dx \tag{9} \\
&= - \int_{-\infty}^0 \int_{-\infty}^y f_X(x) dx dy + \int_0^\infty \int_y^\infty f_X(x) dx dy \tag{10} \\
&= - \int_\infty^0 \mathbb{P}(X \leq y) dy + \int_0^\infty \mathbb{P}(X \geq y) dy \tag{11} \\
&= \int_0^\infty (1 - F_X(y)) dy - \int_{-\infty}^0 F_X(y) dy \tag{12}
\end{align}$$

The result follows by imposing $\mathbb{P}(X > 0) = 1$, which implies $\int_{-\infty}^0 F_X(y) dy = 0$.

Exercise 4.7.8. Prove Theorem 4.16.

Let $X_1, \dots, X_n$ be IID and let $\mu = \mathbb{E}(X_i)$, $\sigma^2 = \mathbb{V}(X_i)$. Then

$$\mathbb{E}\left(\overline{X}_n\right) = \mu, \quad \mathbb{V}\left(\overline{X}_n\right) = \frac{\sigma^2}{n}, \quad \text{and} \quad \mathbb{E}\left(S_n^2\right) = \sigma^2$$

Solution.

For the expected value of sample mean:

$$\mathbb{E}\left(\overline{X}_n\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} n\mu = \mu$$

For the variance of sample mean:

$$\mathbb{V}\left(\overline{X}_n\right) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

For the expected value of sample variance:

$$\begin{aligned}
\mathbb{E}(S_n^2) &= \mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2\right) && (13)\\
&= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2\right) && (14)\\
&= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^n X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2\right) && (15)\\
&= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^n X_i^2 - 2\overline{X}_n\sum_{i=1}^n X_i + n\overline{X}_n^2\right) && (16)\\
&= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^n X_i^2 - 2\overline{X}_n \cdot n\overline{X}_n + n\overline{X}_n^2\right) && (17)\\
&= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^n X_i^2 - n\overline{X}_n^2\right) && (18)\\
&= \frac{1}{n-1}\left(\sum_{i=1}^n \mathbb{E}(X_i^2) - n\mathbb{E}\left(\overline{X}_n^2\right)\right) && (19)\\
&= \frac{1}{n-1}\left(\sum_{i=1}^n \left(\mathbb{V}(X_i) + \mathbb{E}(X_i)^2\right) - n\left(\mathbb{V}\left(\overline{X}_n\right) + \mathbb{E}\left(\overline{X}_n\right)^2\right)\right) && (20)\\
&= \frac{1}{n-1}\left(n\left(\sigma^2+\mu^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) && (21)\\
&= \sigma^2 && (22)
\end{aligned}$$

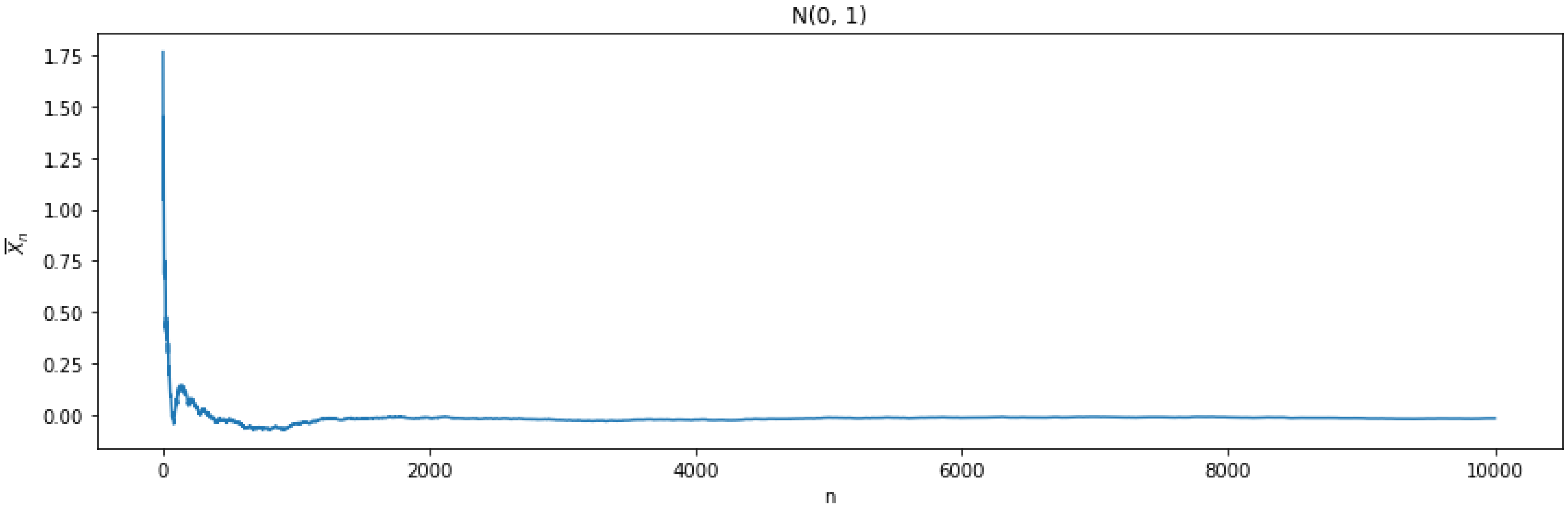
Exercise 4.7.9 (Computer Experiment). Let $X_1, \ldots, X_n$ be $N(0,1)$ random variables and let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$. Plot $\overline{X}_n$ versus n for $n = 1, \ldots, 10,000$. Repeat for $X_1, \ldots, X_n \sim \text{Cauchy}$. Explain why there is such a difference.

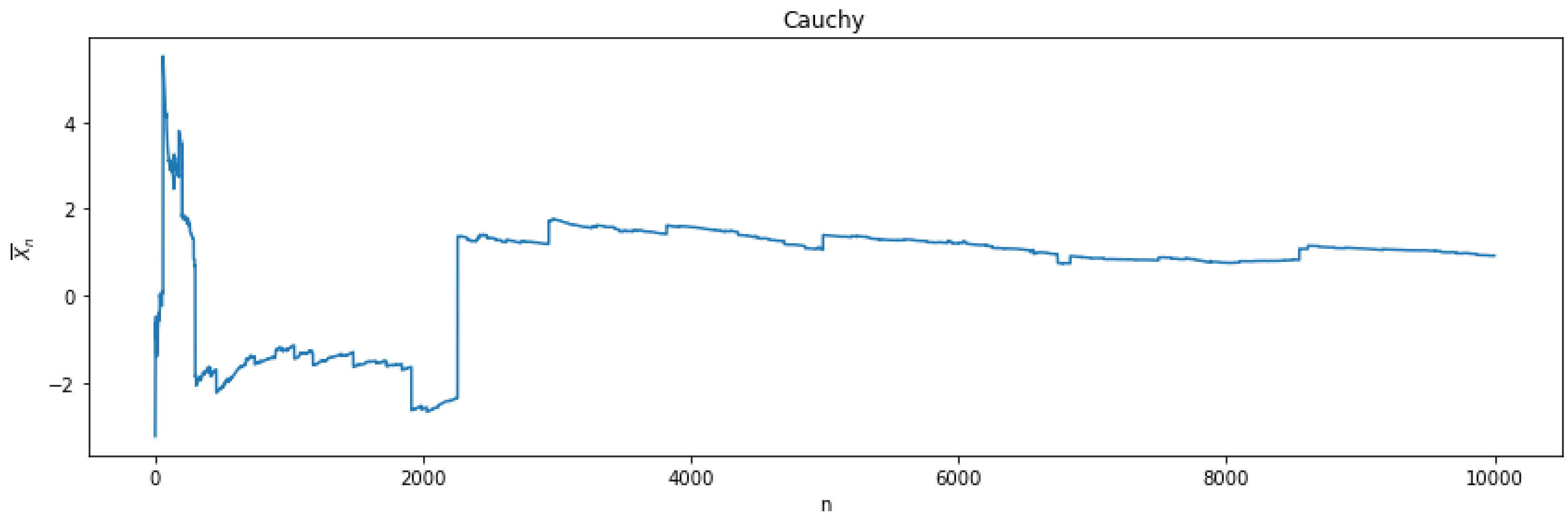
In [1]:

```
import numpy as np
from scipy.stats import norm, cauchy

np.random.seed(0)

N = 10000
X = norm.rvs(size=N)
Y = cauchy.rvs(size = N)
```

In [2]:

```
import matplotlib.pyplot as plt
%matplotlib inline

nn = np.arange(1, N + 1)

plt.figure(figsize=(12, 8))

ax = plt.subplot(2, 1, 1)
ax.plot(nn, np.cumsum(X) / nn)
ax.set_title('N(0, 1)')
ax.set_xlabel('n')
ax.set_ylabel(r'$\overline{X}_n$')

ax = plt.subplot(2, 1, 2)
ax.plot(nn, np.cumsum(Y) / nn)
ax.set_title('Cauchy')
ax.set_xlabel('n')
ax.set_ylabel(r'$\overline{X}_n$')

plt.tight_layout()
plt.show()
```





The mean on the Cauchy distribution is famously undefined: $\overline{X}_n$ is not going to converge.

Exercise 4.7.10. Let $X \sim N(0, 1)$ and let $Y = e^X$. Find $\mathbb{E}(Y)$ and $\mathbb{V}(Y)$.

Solution.

The CDF of Y is, for $y > 0$:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \leq \log y) = \Phi(\log y)$$

and so the PDF is

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}\Phi(\log y) = \frac{d\Phi(\log y)}{d\log y}\frac{d\log y}{dy} = \frac{\phi(\log y)}{y}$$

The expected value is

$$\mathbb{E}(Y) = \int y f_Y(y) dy = \int_0^\infty y \frac{\phi(\log y)}{y} dy = \int_0^\infty \phi(\log y)\; dy = \sqrt{e}$$

The expected value of Y^2 is

$$\mathbb{E}(Y^2) = \int y^2 f_Y(y) dy = \int_0^\infty y^2 \frac{\phi(\log y)}{y} dy = \int_0^\infty y\phi(\log y)\; dy = e^2$$

and so the variance is

$$\mathbb{V}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = e(e-1)$$

Exercise 4.7.11 (Computer Experiment: Simulating the Stock Market). Let $Y_1, Y_2, \dots$ be independent random variables such that $\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = 1/2$. Let $X_n = \sum_{i=1}^n Y_i$. Think of $Y_i = 1$ as "the stock price increased by one dollar" $Y_i = -1$ as "the stock price decreased by one dollar" and X_n as the value of the stock on day n.

(a) Find $\mathbb{E}(X_n)$ and $\mathbb{V}(X_n)$.

(b) Simulate X_n and plot X_n versus n for $n = 1, 2, \dots, 10,000$. Repeat the whole simulation several times. Notice two things. First, it's easy to "see" patterns in the sequence even though it is random. Second, you will find that the runs look very different even though they were generated the same way. How do the calculations in (a) explain the second observation?

Solution.

(a) We have:

$$\mathbb{E}(X_n) = \mathbb{E}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathbb{E}(Y_i) = 0$$

and

$$
\begin{aligned}
\mathbb{E}(X_n^2) &= \mathbb{E}\left(\left(\sum_{i=1}^n Y_i\right)^2\right) && (23)\\
&= \mathbb{E}\left(\sum_{i=1}^n {Y_i}^2 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n Y_i Y_j\right) && (24)\\
&= \sum_{i=1}^n \mathbb{E}({Y_i}^2) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \mathbb{E}(Y_i Y_j) && (25)\\
&= \sum_{i=1}^n 1 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n 0 && (26)\\
&= n && (27)
\end{aligned}
$$

so

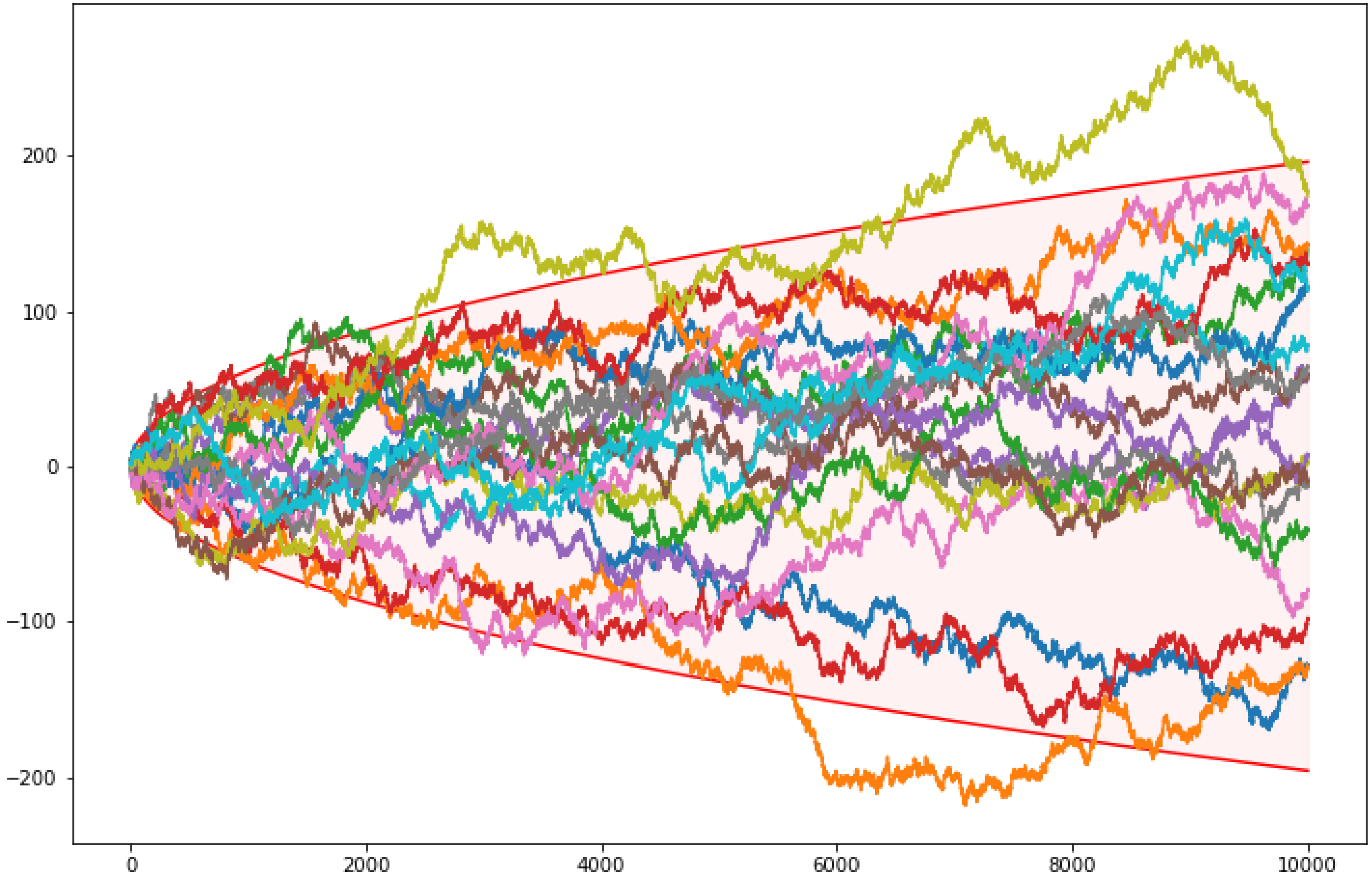
$$\mathbb{V}(X_n) = \mathbb{E}(X_n^2) - \mathbb{E}(X_n)^2 = n$$

(b)

In [3]:
```
import numpy as np
from scipy.stats import norm, bernoulli

N = 10000
B = 20

Y = 2 * bernoulli.rvs(p=1/2, loc=0, size=(B, N), random_state=0) - 1
X = np.cumsum(Y, axis=1)
```

In [4]:
```
import matplotlib.pyplot as plt
%matplotlib inline

plt.figure(figsize=(12, 8))

nn = np.arange(1, N + 1)

z = norm.ppf(0.975)
plt.plot(nn, z * np.sqrt(nn), color='red')
plt.plot(nn, -z * np.sqrt(nn), color='red')
```

```
plt.fill_between(nn, z * np.sqrt(nn), -z * np.sqrt(nn), color='red', alpha=0.05)

for b in range(B):
    plt.plot(nn, X[b])

plt.show()
```



The standard deviation is $\sqrt{n}$ -- it scales up with the square root of the "time". The plot above draws $z_{\alpha/2}\sqrt{n}$ curves -- confidence bands for $1-\alpha = 95\%$ -- that contain most of the randomly generated path.

Exercise 4.7.12. Prove the formulas given in the table at the beginning of Section 4.4 for the Bernoulli, Poisson, Uniform, Exponential, Gamma, and Beta. Here are some hints. For the mean of the Poisson, use the fact that $e^a = \sum_{x=0}^{a} a^x/x!$. To compute the variance, first compute $\mathbb{E}(X(X-1))$.

For the mean of the Gamma, it will help to multiply and divide by $\Gamma(\alpha+1)/\beta^{\alpha+1}$ and use the fact that a Gamma density integrates to 1. For the Beta, multiply and divide by $\Gamma(\alpha+1)\Gamma(\beta)/\Gamma(\alpha+\beta+1)$.

Solution.

We will do all expressions in the table instead (other than multinomial and multivariate normal, where proofs are already provided in the book).

Point mass at p. Let X have a point mass at p. Then:

- $\mathbb{E}(X) = p \cdot 1 = p$
- $\mathbb{E}(X^2) = p^2 \cdot 1 = p^2$
- $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p^2 - p^2 = 0$

Bernoulli. Let $X \sim \text{Bernoulli}(p)$. Then:

- $\mathbb{E}(X) = 1 \cdot p + 0 \cdot (1-p) = p$
- $\mathbb{E}(X^2) = 1 \cdot p + 0 \cdot (1-p) = p$
- $\mathbb{V}(X^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p(1-p)$

Binomial. Let $X \sim \text{Binomial}(n, p)$. Then $X = \sum_{i=1}^n Y_i$, where $Y_i \sim \text{Bernoulli}(p)$ are IID random variables.

- $\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathbb{E}(Y_i) = np$
- $\mathbb{V}(X) = \mathbb{V}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathbb{V}(Y_i) = np(1-p)$

Geometric. Let $X \sim \text{Geometric}(p)$. Then:

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^\infty kp(1-p)^{k-1} && (28) \\
&= \sum_{k=1}^\infty p(1-p)^{k-1} + \sum_{k=2}^\infty (k-1)p(1-p)^{k-1} && (29) \\
&= p\left(1 + (1-p) + (1-p)^2 + \dots\right) + \sum_{k=1}^\infty kp(1-p)^k && (30) \\
&= p\left(\frac{1}{1-(1-p)}\right) + (1-p)\sum_{k=1}^\infty kp(1-p)^{k-1} && (31) \\
&= 1 + (1-p)\mathbb{E}(X) && (32)
\end{aligned}
$$

Solving for the expectation, we get $\mathbb{E}(X) = 1/p$.

We also have:

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} && (33)\\
&= \sum_{k=1}^{\infty} kp(1-p)^{k-1} + \sum_{k=2}^{\infty} (k^2-k)p(1-p)^{k-1} && (34)\\
&= \mathbb{E}(X) + (1-p)\sum_{k=1}^{\infty} (k^2+k)p(1-p)^{k-1} && (35)\\
&= \mathbb{E}(X) + (1-p)\mathbb{E}(X) + (1-p)\sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} && (36)\\
&= \frac{2-p}{p} + (1-p)\mathbb{E}(X^2) && (37)
\end{aligned}$$

Solving for the expectation, we get $\mathbb{E}(X^2) = (2-p)/p^2$.

Finally,

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

Poisson. Let $X \sim \text{Poisson}(\lambda)$. Then:

- $\mathbb{E}(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$
- $\mathbb{E}(X^2) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \lambda \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k e^{-\lambda}}{k!} = \lambda \mathbb{E}(X+1) = \lambda(\lambda+1)$
- $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

Uniform. Let $X \sim \text{Uniform}(a, b)$. Then:

- $\mathbb{E}(X) = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$
- $\mathbb{E}(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{a^2+ab+b}{3}$

- $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{a^2+ab+b}{3} - \frac{a^2+2ab+b^2}{4} = \frac{(b-a)^2}{12}$

Normal. Let $X \sim N(\mu, \sigma^2)$. Converting into a standard normal, we get $Z = (X - \mu)/\sigma \sim N(0, 1)$. Then:

- $\mathbb{E}(X) = \mathbb{E}(\sigma Z + \mu) = \sigma \mathbb{E}(Z) + \mu = \mu$
- $\mathbb{V}(X) = \mathbb{V}(\sigma Z + \mu) = \sigma^2 \mathbb{V}(Z) = \sigma^2$

To prove that the expected value Z is 0, note that the PDF of Z is even, $\phi(z) = \phi(-z)$, so

$$\mathbb{E}(Z) = \int_{-\infty}^{\infty} z\phi(z)dz = \int_{-\infty}^{0} z\phi(z)dz + \int_{0}^{\infty} z\phi(z)dz = \int_{0}^{\infty} -z\phi(-z)dz + \int_{0}^{\infty} z\phi(z)dz = \int_{0}^{\infty} (-z+z)\phi(z) = 0$$

To prove that the variance of Z is 0, write out the integral explicitly for the expectation of Z^2,

$$\mathbb{E}(Z^2) = \int_{-\infty}^{\infty} z^2\phi(z)dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz = \left[\Phi(z) - \frac{1}{\sqrt{2\pi}} z e^{-z^2/2} \right]_{-\infty}^{\infty} = \lim_{x \rightarrow +\infty} \Phi(x) - \lim_{x \rightarrow -\infty} \Phi(x) = 1 - 0 = 1$$

and so

$$\mathbb{V}(Z) = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = 1 - 0 = 1$$

Exponential. Let $X \sim \text{Exponential}(\beta)$. Then:

- $\mathbb{E}(X) = \int_0^\infty x \frac{1}{\beta} e^{-x/\beta} dx = \frac{1}{\beta} \int_0^\infty x e^{-x/\beta} dx = \frac{1}{\beta}\beta^2 = \beta$
- $\mathbb{E}(X^2) = \int_0^\infty x^2 \frac{1}{\beta} e^{-x/\beta} dx = \frac{1}{\beta} \int_0^\infty x^2 e^{-x/\beta} dx = \frac{1}{\beta} 2\beta^3 = 2\beta^2$
- $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 2\beta^2 - \beta^2 = \beta^2$

Gamma. Let $X \sim \text{Gamma}(\alpha, \beta)$. The PDF is

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\beta x} \quad \text{for } x > 0$$

We have:

$$
\begin{aligned}
\mathbb{E}(X) &= \int x f_X(x) dx && (38) \\
&= \int_0^\infty x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx && (39) \\
&= \frac{\alpha}{\beta} \int_0^\infty \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^{\alpha} e^{-\beta x} dx && (40) \\
&= \frac{\alpha}{\beta} && (41)
\end{aligned}
$$

where we used that

- $\alpha\Gamma(\alpha) = \Gamma(\alpha+1)$,
- and last integral is the PDF of $\text{Gamma}(\alpha+1, \beta)$, integrated over its entire domain.

We also have:

$$
\begin{aligned}
\mathbb{E}(X^2) &= \int x^2 f_X(x) dx && (42) \\
&= \int_0^\infty x^2 \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx && (43) \\
&= \frac{\alpha(\alpha+1)}{\beta^2} \int_0^\infty \frac{\beta^{\alpha+2}}{\Gamma(\alpha+2)} x^{\alpha+1} e^{-\beta x} dx && (44) \\
&= \frac{\alpha(\alpha+1)}{\beta^2} && (45)
\end{aligned}
$$

- $\alpha(\alpha+1)\Gamma(\alpha+1) = \Gamma(\alpha+2)$,
- and last integral is the PDF of $\text{Gamma}(\alpha+2, \beta)$, integrated over its entire domain.

Therefore,

$$
\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}
$$

Beta. Let $X \sim \text{Beta}(\alpha, \beta)$. The PDF is

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad \text{for } x > 0$$

We have:

$$\begin{align}
\mathbb{E}(X) &= \int x f_X(x) dx \tag{46}\\
&= \int_0^\infty x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx \tag{47}\\
&= \frac{\alpha}{\alpha+\beta} \int_0^\infty \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{\alpha}(1-x)^{\beta-1} dx \tag{48}\\
&= \frac{\alpha}{\alpha+\beta} \tag{49}
\end{align}$$

where we used that

- $\alpha\Gamma(\alpha) = \Gamma(\alpha+1)$,
- $(\alpha+\beta)\Gamma(\alpha+\beta) = \Gamma(\alpha+\beta+1)$,
- and the last integral is the PDF of $\text{Beta}(\alpha+1, \beta)$, integrated over its entire domain.

We also have:

$$\begin{align}
\mathbb{E}(X^2) &= \int x^2 f_X(x) dx \tag{50}\\
&= \int_0^\infty x^2 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx \tag{51}\\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \int_0^\infty \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+2)\Gamma(\beta)} x^{\alpha+1}(1-x)^{\beta-1} dx \tag{52}\\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \tag{53}
\end{align}$$

where we used that

- $\alpha(\alpha+1)\Gamma(\alpha+1) = \Gamma(\alpha+2)$,
- $(\alpha+\beta)(\alpha+\beta+1)\Gamma(\alpha+\beta+1) = \Gamma(\alpha+\beta+2)$,

- and the last integral is the PDF of $\text{Beta}(\alpha + 2, \beta)$, integrated over its entire domain.

Therefore,

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

t-student. Let $X \sim t_\nu$. The PDF for the t-student distribution is

$$f_X(x) = \frac{1}{\sqrt{\upsilon\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{\nu}\right)^{(\nu+1)/2}}$$

Since the PDF is even, $f_X(x) = f_X(-x)$, the expectation will be 0 when it is defined:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{0} x f_X(x) dx + \int_{0}^{\infty} x f_X(x) dx = \int_{0}^{\infty} -x f_X(-x) dx + \int_{0}^{\infty} x f_X(x) dx = \int_{0}^{\infty} (-x + x) f_X(x) dx = 0$$

But

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \frac{1}{\sqrt{\upsilon\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^{\infty} x \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} dx$$

For the expectation of X^2, assuming it is defined, we have:

$$
\begin{aligned}
\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x) dx && (54) \\
&= \frac{1}{\sqrt{\upsilon\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^{\infty} x^2 \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} dx && (55) \\
&= \frac{\nu}{\sqrt{\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \int_0^1 y^{\nu/2-2} (1-y)^{1/2} dy && (56) \\
&= \frac{\nu}{\sqrt{\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{\Gamma\left(\frac{\nu}{2}-1\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)} && (57) \\
&= \frac{\nu}{\nu-2} && (58)
\end{aligned}
$$

where we used:

- A variable replacement $y = \left(1 + \frac{x^2}{\nu}\right)^{-1}$
- The property that $\int_0^1 y^{p-1}(1-y)^{q-1} dy = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$, since this is the integral of the PDF of $\Gamma(p, q)$ scaled by a factor of $\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$, with $p = \nu/2 - 1, q = 3/2$
- $\Gamma(3/2) = \sqrt{\pi}$

Finally,

$$
\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{\nu}{\nu-2}
$$

Reference: https://math.stackexchange.com/a/1502519

χ^2 **distribution**. Let $X \sim \chi^2_k$. Then X has the same distributions as the sum of squares of k IID standard Normal random variables, $X = \sum_{i=1}^k Z_i^2$, $Z_i \sim N(0, 1)$.

The expectation of X can then be computed:

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^k Z_i^2\right) = \sum_{i=1}^k \mathbb{E}(Z_i^2) = \sum_{i=1}^k (\mathbb{V}(Z_i) + \mathbb{E}(Z_i)^2) = \sum_{i=1}^k (1+0) = k$$

The expectation of X^2 is:

$$\begin{aligned}
\mathbb{E}(X^2) &= \mathbb{E}\left(\left(\sum_{i=1}^k Z_i^2\right)^2\right) && (59)\\
&= \mathbb{E}\left(\sum_{i=1}^k Z_i^4 + \sum_{i=1}^k \sum_{j=1;j\neq i}^k Z_i^2 Z_j^2\right) && (60)\\
&= \sum_{i=1}^k \mathbb{E}(Z_i^4) + \sum_{i=1}^k \sum_{j=1;j\neq i}^k \mathbb{E}(Z_i^2)\mathbb{E}(Z_j^2) && (61)
\end{aligned}$$

But we have:

$$\mathbb{E}(Z_i^2) = \mathbb{V}(Z_i) + \mathbb{E}(Z_i)^2 = 1 + 0 = 1$$

and, using moment generating functions,

$$M_Z(t) = e^{t^2/2}$$

and taking the fourth derivative,

$$M_Z^{(4)}(t) = 3M_Z^{(2)}(t) + tM_Z^{(3)}(t)$$

Setting $t = 0$ gives us $\mathbb{E}(Z_i^4) = 3$.

Replacing it back on the expectation expression for X^2,

$$\mathbb{E}(X^2) = \sum_{i=1}^k 3 + \sum_{i=1}^k \sum_{j=1;j\neq i}^k 1 \cdot 1 = 3k + k(k-1) = k^2 + 2k$$

Therefore,

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = k^2 + 2k - k^2 = 2k$$

The proofs for the multinomial and mutivariate normal distribution expressions are provided in the book text (and there are notes above).

Exercise 4.7.13. Suppose we generate a random variable X in the following way. First we flip a fair coin. If the coin is heads, take X to have a $\text{Uniform}(0, 1)$ distribution. If the coin is tails, take X to have a $\text{Uniform}(3, 4)$ distribution.

(a) Find the mean of X.

(b) Find the standard deviation of X.

Solution. We have $X = CU_1 + (1 - C)U_2$, where $U \sim \text{Bernoulli}(1/2)$, $U_1 \sim \text{Uniform}(0, 1)$ and $U_2 \sim \text{Uniform}(0, 2)$ are all independent.

(a)

$$\mathbb{E}(X) = \mathbb{E}(CU_1 + (1 - C)U_2) = \mathbb{E}(C)\mathbb{E}(U_1) + (1 - \mathbb{E}(C))\mathbb{E}(U_2) = \frac{1}{2}\left(\frac{1}{2} + \frac{7}{2}\right) = 2$$

(b)

$$X^2 = (CU_1 + (1 - C)U_2)^2 = C^2U_1^2 + (1 - C)^2U_2^2 + 2C(1 - C)U_1U_2 = C^2U_1^2 + (1 - C)^2U_2^2$$

so

$$\begin{aligned}
\mathbb{E}(X^2) &= \mathbb{E}(C^2)\mathbb{E}(U_1^2) + \mathbb{E}((1 - C)^2)\mathbb{E}(U_2^2) && (62)\\
&= \mathbb{E}(C)\mathbb{E}(U_1^2) + \mathbb{E}(1 - C)\mathbb{E}(U_2^2) && (63)\\
&= \frac{1}{2}\left(\frac{1}{3} + \frac{37}{3}\right) = \frac{19}{3} && (64)
\end{aligned}$$

and then

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{19}{3} - 2^2 = \frac{7}{3}$$

and so the standard deviation is $\sqrt{\mathbb{V}(X)} = \sqrt{7/3}$.

Exercise 4.17.14. Let $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ be random variables and let $a_1, \dots, a_m$ and $b_1, \dots, b_n$ be constants. Show that

$$\text{Cov}\left(\sum_{i=1}^m a_iX_i, \sum_{j=1}^n b_jY_j\right) = \sum_{i=1}^m \sum_{j=1}^n a_ib_j\text{Cov}(X_i, Y_j)$$

Solution. We have:

$$\begin{aligned}
\text{Cov}\left(\sum_{i=1}^m a_i X_i, Y\right) &= \mathbb{E}\left(\left(\sum_{i=1}^m a_i X_i\right) Y\right) - \mathbb{E}\left(\sum_{i=1}^m a_i X_i\right)\mathbb{E}(Y) && (65)\\
&= \sum_{i=1}^m \mathbb{E}(a_i X_i Y) - \left(\sum_{i=1}^m a_i \mathbb{E}(X_i)\right)\mathbb{E}(Y) && (66)\\
&= \sum_{i=1}^m \mathbb{E}(a_i X_i Y) - a_i \mathbb{E}(X_i)\mathbb{E}(Y) && (67)\\
&= \sum_{i=1}^m a_i \text{Cov}(X_i, Y) && (68)
\end{aligned}$$

and, since $\text{Cov}(A, B) = \text{Cov}(B, A)$,

$$\text{Cov}\left(X, \sum_{j=1}^n b_j Y_j\right) = \sum_{j=1}^n b_i \text{Cov}(X, Y_j)$$

Applying this for each X_i, we get the result.

Exercise 4.17.15. Let

$$f_{X,Y} = \begin{cases} \frac{1}{3}(x+y) & \text{if } 0 \le x \le 1, 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Find $\mathbb{V}(2X - 3Y + 8)$.

Solution. Let $r(x, y) = 2x - 3y$. Then:

$$\mathbb{V}(2X - 3Y + 8) = \mathbb{V}(2X - 3Y) = \mathbb{V}(r(X, Y))$$

Calculating the expectation of $r(X, Y)$ and $r(X, Y)^2$:

$$\mathbb{E}(r(X, Y)) = \int_0^1 \int_0^2 r(x, y) f(x, y) dy dx = \int_0^1 \int_0^2 \frac{1}{3}(2x - 3y)(x + y) dy dx = \int_0^1 \frac{2}{3}(2x^2 - x - 4) dx = -\frac{23}{9}$$

and

$$\mathbb{E}(r(X,Y)^2) = \int_0^1 \int_0^2 r(x,y)^2 f(x,y) dy dx = \int_0^1 \int_0^2 \frac{1}{3}(2x-3y)^2(x+y) dy dx = \int_0^1 \frac{4}{3}(2x^3 - 4x^2 - 2x + 9) dx = \frac{86}{9}$$

and so

$$\mathbb{V}(r(X,Y)) = \mathbb{E}(r(X,Y)^2) - \mathbb{E}(r(X,Y))^2 = \frac{86}{9} - \frac{23^2}{9^2} = \frac{245}{81}$$

Exercise 4.17.16. Let $r(x)$ be a function of x and let $s(y)$ be a function of y. Show that

$$\mathbb{E}(r(X)s(Y)|X) = r(X)\mathbb{E}(s(Y)|X)$$

Also, show that $\mathbb{E}(r(X)|X) = r(X)$.

Solution. We have:

$$\mathbb{E}(r(X)s(Y)|X = x) = \int r(x)s(y)f(x,y)dy = r(x)\int s(y)f(x,y)dy = r(x)\mathbb{E}(s(Y)|X = x)$$

and so the random variable $\mathbb{E}(r(X)s(Y)|X)$ takes the same value as the variable $r(X)\mathbb{E}(s(Y)|X)$ for each $X = x$ -- therefore the random variables are equal.

In particular, when $s(y) = 1$ for all y, we have $\mathbb{E}(r(X)|X) = r(X)$.

Exercise 4.17.17. Prove that

$$\mathbb{V}(Y) = \mathbb{E}\mathbb{V}(Y|X) + \mathbb{V}\mathbb{E}(Y|X)$$

Hint: Let $m = \mathbb{E}(Y)$ and let $b(x) = \mathbb{E}(Y|X = x)$. Note that $\mathbb{E}(b(X)) = \mathbb{E}\mathbb{E}(Y|X) = \mathbb{E}(Y) = m$. Bear in mind that b is a function of x. Now write

$$\mathbb{V}(Y) = \mathbb{E}((Y-m)^2) = \mathbb{E}(((Y - b(X)) + (b(X) - m))^2)$$

Expand the square and take the expectation. You then have to take the expectation of three terms. In each case, use the rule of iterated expectation, i.e. $\mathbb{E}(\text{stuff}) = \mathbb{E}(\mathbb{E}(\text{stuff}|X))$.

Solution. We have:

$$\begin{aligned}
\mathbb{V}(Y) &= \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 && (69)\\
&= \mathbb{E}(\mathbb{E}(Y^2|X)) - \mathbb{E}(\mathbb{E}(Y|X))^2 && (70)\\
&= \mathbb{E}\left(\mathbb{V}(Y|X) + \mathbb{E}(Y|X)^2\right) - \mathbb{E}(\mathbb{E}(Y|X))^2 && (71)\\
&= \mathbb{E}(\mathbb{V}(Y|X)) + \left(\mathbb{E}(\mathbb{E}(Y|X)^2) - \mathbb{E}(\mathbb{E}(Y|X))^2\right) && (72)\\
&= \mathbb{E}(\mathbb{V}(Y|X) + \mathbb{V}(\mathbb{E}(Y|X)) && (73)
\end{aligned}$$

Exercise 4.17.18. Show that if $\mathbb{E}(X|Y=y) = c$ for some constant c then X and Y are uncorrelated.

Solution. We have:

$$\mathbb{E}(XY) = \int \mathbb{E}(XY|Y=y)dF_Y(y) = \int y\mathbb{E}(X|Y=y)dF_Y(y) = \int cy\mathbb{E}(X|Y=y)dF_Y(y) = c\,\mathbb{E}(Y)$$

and

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(c) = c$$

so $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, and so $\text{Cov}(X,Y) = 0$, and so X and Y are uncorrelated.

Exercise 4.17.19. This question is to help you understand the idea of **sampling distribution**. Let $X_1, \dots, X_n$ be IID with mean μ and variance σ^2. Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$. Then $\overline{X}_n$ is a **statistic**, that is, a function of the data. Since $\overline{X}_n$ is a random variable, it has a distribution. This distribution is called the *sampling distribution of the statistic*. Recall from Theorem 4.16 that $\mathbb{E}(\overline{X}_n) = \mu$ and $\mathbb{V}(\overline{X}_n) = \sigma^2/n$. Don't confuse the distribution of the data f_X and the distribution of the statistic $f_{\overline{X}_n}$. To make this clear, let $X_1, \dots, X_n \sim \text{Uniform}(0,1)$. Let f_X be the density of the $\text{Uniform}(0,1)$. Plot f_X. Now let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$. Find $\mathbb{E}(\overline{X}_n)$ and $\mathbb{V}(\overline{X}_n)$. Plot them as a function of n. Comment. Now simulate the distribution of $\overline{X}_n$ for $n = 1, 5, 25, 100$. Check the simulated values of $\mathbb{E}(\overline{X}_n)$ and $\mathbb{V}(\overline{X}_n)$ agree with your theoretical calculations. What do you notice about the sampling distribution of $\overline{X}_n$ as it increases?

Solution.

$$\mathbb{E}\left(\overline{X}_n\right) = \mathbb{E}\left(n^{-1}\sum_{i=1}^n X_i\right) = n^{-1}\sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{2}$$

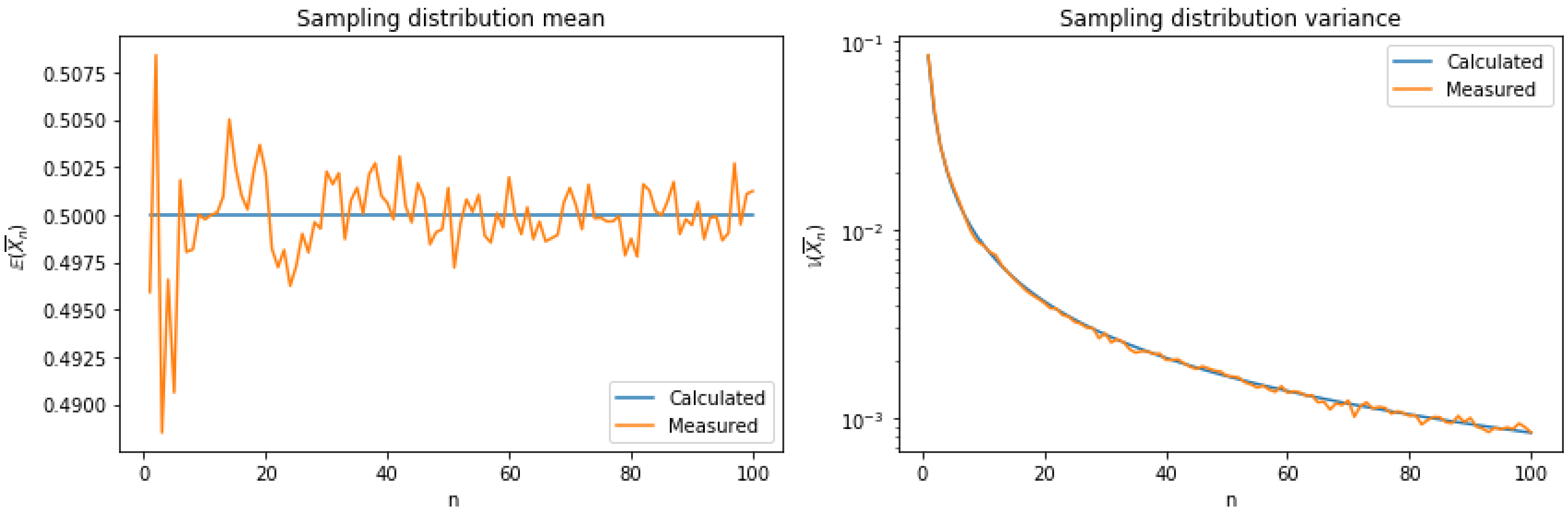
and

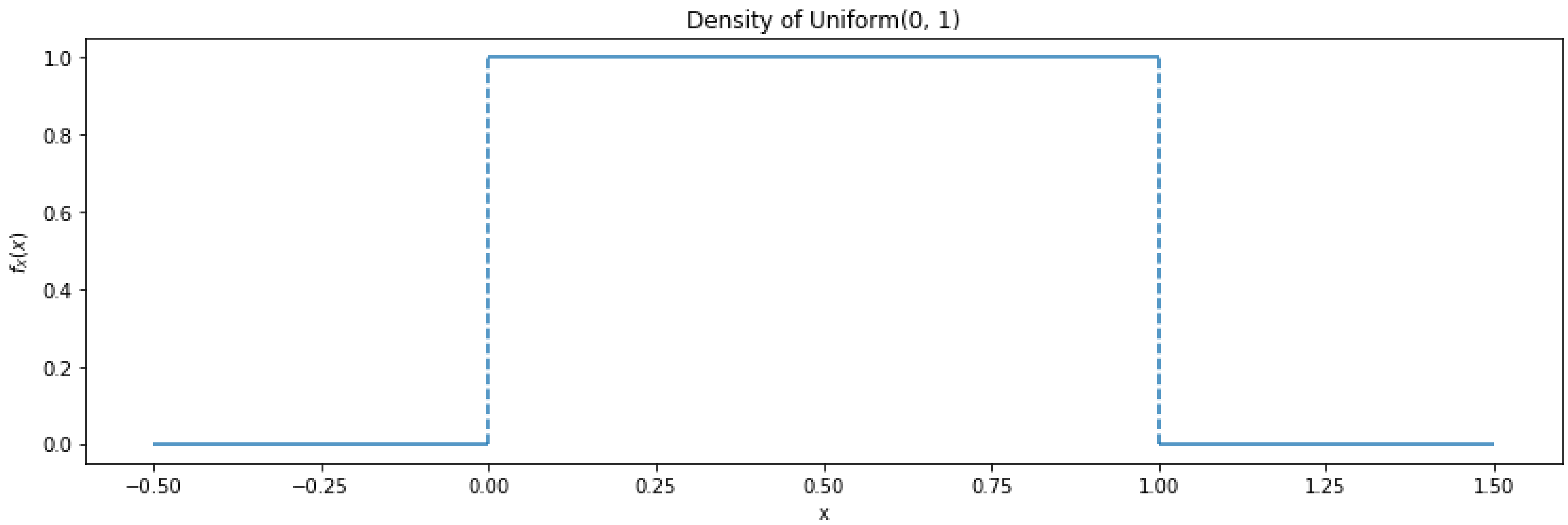
$$\mathbb{V}\left(\overline{X}_n\right) = \mathbb{V}\left(n^{-1}\sum_{i=1}^n X_i\right) = n^{-2}\sum_{i=1}^n \mathbb{V}(X_i) = \frac{1}{12n}$$

In [5]:
```
import numpy as np

np.random.seed(0)

B = 1000

E_overline_X = np.empty(100)
V_overline_X = np.empty(100)

for n in range(1, 101):
    X_n = np.random.uniform(low=0, high=1, size=(B, n)).mean(axis=1)
    E_overline_X[n - 1] = X_n.mean()
    V_overline_X[n - 1] = X_n.var()
```

In [6]:
```
import matplotlib.pyplot as plt
%matplotlib inline

plt.figure(figsize=(12, 8))

ax = plt.subplot(212)
ax.hlines(0, xmin=-0.5, xmax=0, color='C0')
ax.hlines(1, xmin=0, xmax=1, color='C0')
ax.hlines(0, xmin=1, xmax=1.5, color='C0')
ax.vlines([0, 1], ymin=0, ymax=1, color='C0', linestyle='dashed')
ax.set_xlabel('x')
ax.set_ylabel(r'$f_X(x)$')
ax.set_title('Density of Uniform(0, 1)')

nn = np.arange(1, 101)

ax = plt.subplot(221)
ax.plot(nn, 1/2 * np.ones(100), label='Calculated')
ax.plot(nn, E_overline_X, label='Measured')
ax.set_xlabel('n')
ax.set_ylabel(r'$\mathbb{E}(\overline{X}_n)$')
ax.set_title('Sampling distribution mean')
ax.legend(loc='lower right')

ax = plt.subplot(222)
ax.plot(nn, 1 / (12 * nn), label='Calculated')
ax.plot(nn, V_overline_X, label='Measured')
ax.set_xlabel('n')
ax.set_yscale('log')
```

```
ax.set_ylabel(r'$\mathbb{V}(\overline{X}_n)$')
ax.set_title('Sampling distribution variance')
ax.legend(loc='upper right')

plt.tight_layout()
plt.show()
```





Calculated and simulated values agree.

Exercise 4.17.20. Prove Lemma 4.20.

If a is a vector and X is a random vector with mean μ and variance Σ then

$$\mathbb{E}(a^T X) = a^T \mu \quad \text{and} \quad \mathbb{V}(a^T X) = a^T \Sigma a$$

If A is a matrix then

$$\mathbb{E}(AX) = A\mu \quad \text{and} \quad \mathbb{V}(AX) = A\Sigma A^T$$

Solution.

We have:

$$\mathbb{E}(a^T X) = \begin{pmatrix} \mathbb{E}(a_1 X_1) \\ \mathbb{E}(a_2 X_2) \\ \cdots \\ \mathbb{E}(a_k X_k) \end{pmatrix} = \begin{pmatrix} a_1 \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \\ \cdots \\ \mathbb{E}(X_k) \end{pmatrix} = a^T \mu$$

and

$$\mathbb{V}(a^T X) = \mathbb{E}((a^T (X - \mu)(a^T (X - \mu))^T) = \mathbb{E}((a^T (X - \mu)(X - \mu)^T a) = a^T \Sigma a$$

Similarly, for the matrix case,

$$\mathbb{E}(AX) = \begin{pmatrix} \mathbb{E}\left(\sum_{j=1}^k a_{1,j} X_j\right) \\ \mathbb{E}\left(\sum_{j=1}^k a_{2,j} X_j\right) \\ \cdots \\ \mathbb{E}\left(\sum_{j=1}^k a_{k,j} X_j\right) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^k a_{1,j} \mathbb{E}(X_j) \\ \sum_{j=1}^k a_{2,j} \mathbb{E}(X_j) \\ \cdots \\ \sum_{j=1}^k a_{k,j} \mathbb{E}(X_j) \end{pmatrix} = A\mu$$

and

$$\mathbb{V}(AX) = \mathbb{E}((A(X - \mu)(A(X - \mu))^T) = \mathbb{E}((A(X - \mu)(X - \mu)^T A^T) = A\Sigma A^T$$