Self Assasment

Peter Fürjész - Examples

2.1/a - Solution

Using (1), show that the Taylor series of $\exp(x) = \mathrm{e}^x$ around the point zero is given by

$$\mathrm{e}^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

<u>Solution:</u>

$f(x) = e^x \quad and\ a = 0$

for zero degree polynomial approximation about $a = 0$:

$e^x = f(a) = 1 + O(x)$

for first degree polynomial approximation about : $a = 0$:

$e^x = f(a) + f'(a)(x-a) \ = 1 \ + x + O\left(x^2\right)$

for second degree polynomial approximation about: $a = 0$:

$e^x = 1 + 1 \cdot x + \dfrac{f''(a)}{2!} \cdot x^2 = 1 + x + \dfrac{1}{2}x^2 + O\left(x^2\right)$

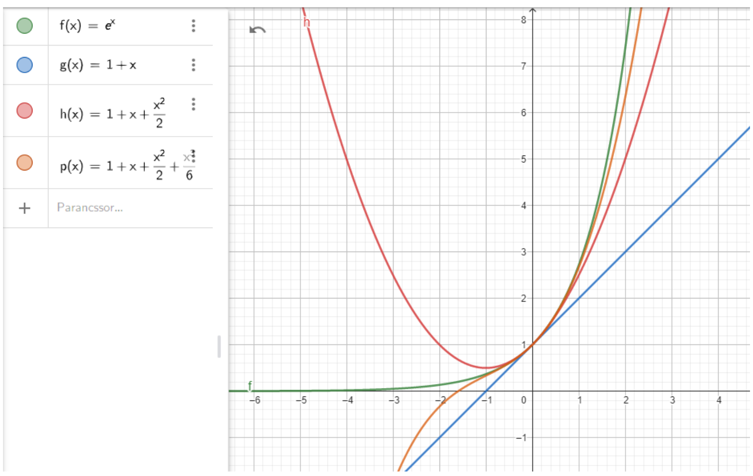
....

n^{th} degree Taylor Polynomial for $f(a) = e^x$ about $a = 0$

$T_n(x) = 1 + x + \dfrac{1}{2!}x^2 + \dfrac{1}{3!}x^3 + \cdots + \dfrac{1}{n!}x^n = \displaystyle\sum_{k=0}^{n} \frac{x^k}{k!}$

Source: Geo Gebra

Figure 1: Approximating the function with polynomials



2.1/b - Solution

Using (2), compute (what turns out to be the expectation of a Poisson random variable)

$$\sum_{y=0}^{\infty} \frac{y\mathrm{e}^{-\lambda}\lambda^y}{y!}, \quad \lambda > 0$$

Solution:

$\mathrm{e}^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ // Multiplying the equation by x

$x * \mathrm{e}^x = \sum_{n=0}^{\infty} \frac{x^{n+1}}{n!}$ // Deriving both sides of the eqution

$e^x + x * e^x = \sum_{n=0}^{\infty}(n+1) * \frac{x^n}{n!} = \sum_{n=0}^{\infty} n * \frac{x^n}{n!} + \sum_{n=0}^{\infty} \frac{x^n}{n!} = \sum_{n=0}^{\infty} n * \frac{x^n}{n!} + \mathrm{e}^x$ // minus e^x

$x * e^x = \sum_{n=0}^{\infty} n * \frac{x^n}{n!}$ // Let $x = \lambda$ and $n = y$ and multiplying with $e^{-\lambda}$

$\lambda = \sum_{y=0}^{\infty} \frac{y\mathrm{e}^{-\lambda}\lambda^y}{y!}$

2.2/a - Solution

For $a = 0$ and $I = (-1, 1)$, compute the Taylor series of $f = \log(1 + x)$ for $x \in I$ to show that

$$\log(1 + x) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{x^k}{k}, \quad x \in (-1, 1)$$

Solution:

$$\begin{aligned}
&f(x) = f(0) + xf'(0) + \tfrac{x^2}{2!}f''(0) + \tfrac{x^3}{3!}f'''(0) + \tfrac{x^4}{4!}f''(0) + \ldots \\
&f(x) = \log(1 + x) && f(0) = 0 \\
&f'(x) = \tfrac{1}{1+x} && f'(0) = 1 \\
&f''(x) = \tfrac{-1}{(1+x)^2} = -(1 + x)^{-2} && f''(0) = -1 \\
&f'''(x) = -(-2)(1 + x)^{-3} && f'''(0) = 2 \\
&f^{IV}(x) = 2(-3) \cdot (1 + x)^{-4} && f''(0) = -2 \cdot 3 \\
&f^{V}(x) = (-2)3 \cdot (-4)(1 + x)^{-5} && f^{v}(0) = 2 \cdot 3 \cdot 4
\end{aligned} \tag{1}$$

$$\begin{aligned}
\log(1 + x) &= x - \frac{x^2}{2} + \frac{2x^3}{6} - \frac{6x^4}{24} + \frac{24x^5}{120} - \ldots \\
&= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \ldots\ldots
\end{aligned}$$

Substituting $x = 1$:

$$\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots .$$

2.2/b - Solution

Compute the Taylor series of $g(x) = \log$ about the point $a = 1$, and confirm:

Taking $I = (-1, 1]$, so that:

$$\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

Solution:

Knowing that:

$$f(x) = \ln x \to f(1) = 0$$

$$f'(x) = 1/x \to f'(1) = 1$$

$$f''(x) = -\frac{1}{x^2} \to f''(1) = -1$$

$$f'''(x) = \frac{2}{x^3} \to f'''(1) = 2$$

$$f^{(4)}(x) = -\frac{(2 \cdot 3)}{x^4} \to f^{(4)}(1) = -6$$

$$f^{(5)}(x) = \frac{(2 \cdot 3 \cdot 4)}{x^5} \to f^{(5)}(1) = 24$$

Substituting into the taylor formula:

$$\begin{aligned}
&\frac{0}{0!}(x-1)^0 + \frac{1}{1!}(x-1) - \frac{1}{2!}(x-1)^2 + \frac{2}{3!}(x-1)^3 - \frac{6}{4!}(x-1)^4 + \frac{24}{5!}(x-1)^5 - \dots \\
=&(x-1) - \frac{1!}{2!}(x-1)^2 + \frac{2!}{3!}(x-1)^3 - \frac{3!}{4!}(x-1)^4 + \frac{4!}{5!}(x-1)^5 - \dots \\
=&(x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \frac{1}{4}(x-1)^4 + \frac{1}{5}(x-1)^5 - \dots
\end{aligned} \tag{2}$$

2.3/a - Solution

(Understanding of existence of integral, Cauchy principle value) // Let $f(x) = x/(1+x^2)$. For $K > 0$ derive

$$\int_0^K \frac{x}{1+x^2}\,\mathrm{d}x \text{ and } \int_{-K}^0 \frac{x}{1+x^2}\,\mathrm{d}x, \text{ and thus } \int_{-K}^K \frac{x}{1+x^2}\,\mathrm{d}x.$$

Let:

$$\begin{aligned} u &= 1 + x^2 \\ du &= 2xdx \\ \frac{du}{2} &= xdx \end{aligned} \tag{3}$$

Then:

$$\int \frac{x}{1+x^2}dx = \int \frac{1}{u} \cdot \frac{du}{2} = \frac{1}{2}\int \frac{1}{u}du = \frac{1}{2}\ln|u| = \frac{1}{2}\ln\left|1+x^2\right| + C \tag{4}$$

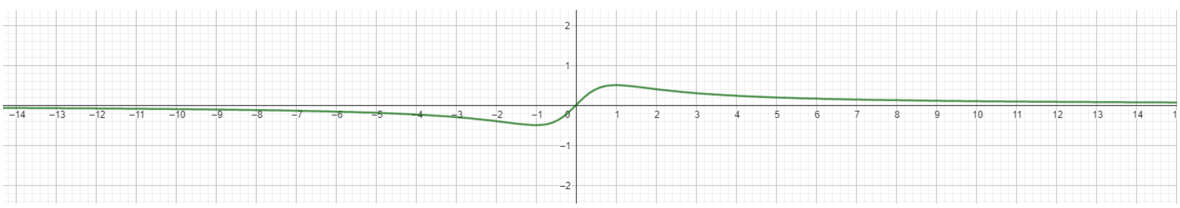
2.3/b - Solution

What can you conclude about

$$\int_{-\infty}^{\infty} \frac{x}{1+x^2}\,\mathrm{d}x?$$

Source: Geo Gebra

Figure 2: Visualisation of the function



does not exist, although for any given b > 0 we get:

$$\int_{-b}^{b} \frac{2x}{1+x^2}dx = 0$$

because the integrand is an odd function. However

$$\int_0^b \frac{x}{1+x^2}dx = \frac{1}{2}\ln\left(1+b^2\right):$$

and its limit is $+\infty$, when $b \to \infty$ and the limit of the integral on $(-\infty, 0]$ is $-\infty$, so the sum does not exist.

2.4 - Solution

State the definition the binomial coefficient (i.e., " n choose k "):

- $\begin{pmatrix} n \\ k \end{pmatrix}$ is the number of ways to select k objects from a set of n objects.
- $\begin{pmatrix} n \\ k \end{pmatrix}$ is the coefficient of $x^{n-k}y^k$ in the expansion of $(x+y)^n$

Explain in words why:

The triangle is symmetric. In any row, entries on the left side are mirrored on the right side.

and algebraically prove, that:

$$\begin{pmatrix} n \\ k \end{pmatrix} = \begin{pmatrix} n \\ n-k \end{pmatrix}$$

By definition:

$$\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!} \tag{5}$$

$$\begin{pmatrix} n \\ n-k \end{pmatrix} = \frac{n!}{(n-k)!(n-(n-k))!} \tag{6}$$

Where:

$$n - \ (n-k) = n + (-1)(n-k) = n + (-1)n - (-1)k = n - n + k = k \tag{7}$$

Then:

$$\frac{n!}{(n-k)!(n-(n-k))!} = \frac{n!}{(n-k)!k!} = \begin{pmatrix} n \\ k \end{pmatrix} \tag{8}$$

2.5/a - Solution

(State the binomial theorem. (Don't forget to indicate the sets on which the variables you define are defined.)

Let x and y be variables and n a natural number, then

$$(x+y)^n = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} x^{n-k} y^k \tag{9}$$

2.5/b - Solution

State the basic calculus definition of the first derivative of a function (the Newton quotient) in terms of a limit.

We say that f is differentiable at x_0 if the following limi exists and it is finite:

$$\lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$$

This limit is called the derivative of f at the point x_0, its notation is $f'(x_0)$, We say that the function f is differentiable in an interval, if it is differentiable at every interior point of the interval. The quotient above is called the difference quotient of f at the point x_0.

2.5/c - Solution

(c) Use the definition of derivative to derive the derivative of $f(x) = x^n$, where $n \in \{1, 2, 3, \ldots\}$ is a natural number.

$$\frac{d}{dx}(x^n) = \lim_{h\to 0} \frac{(x+h)^n - x^n}{h} \tag{10}$$

By expanding the numerator, we obtain that the numerator equals:

$$n^X + \begin{pmatrix} n \\ 1 \end{pmatrix} x^{n-1}h + \ldots + h^n - x^n \tag{11}$$

So the numerator equals:

$$\begin{pmatrix} n \\ 1 \end{pmatrix} x^{n-1}h + \ldots + h^{n-1}x \begin{pmatrix} n \\ n-1 \end{pmatrix} + h^n \tag{12}$$

The denominator of the fraction is h, so we divide by h.

$$\begin{pmatrix} n \\ 1 \end{pmatrix} x^{n-1} + \begin{pmatrix} n \\ 2 \end{pmatrix} x^{n-2}h + \ldots + h^{n-1} \tag{13}$$

The derivative then becomes the limit, as h approaches 0 .We plug in $h = 0$, so the expression becomes nx^{n-1}. So:

$$\frac{d}{dx}\left(x^n\right) = nx^{n-1} \tag{14}$$

2.6 - Solution

For a geometric series $\left\{ar^k\right\}_{k=0}^{\infty}$, $a \in R$ and $|r| < 1$, prove that the infinite sum has the following closed form

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$$

$$\begin{aligned} S_n &= a_1 + a_2 + a_3 + a_4 + \ldots a_{n-1} + a_n \\ S_n &= a_1 + a_1 r + a_1 r^2 + a_1 r^3 + \ldots a_1 (r)^{n-2} + a_1 r^{n-1} \\ rS_n &= a_1 r + a_1 r^2 + a_1 r^3 + a_1 r^4 + \ldots a_1 (r)^{n-1} + a_1 r^n \end{aligned} \tag{15}$$

Multiply equation III by -1 and sum equations II and III. , we got :

$$S_n - rS_n = a_1 - a_1 r^n \tag{16}$$

By factoring out: S_n and a_1 and devideing by $1 - r$ we got:

$$S_n = \frac{a_1\left[1 - r^n\right]}{1 - r} \tag{17}$$

If n tends to ∞, then $1 - r^n$ tends to 1, so

$$S_\infty = \frac{a_1}{1 - r} \tag{18}$$

2.7 - Solution

Evaluate the integral:

$$\int_{-\infty}^{\infty} x^2 \mathrm{e}^{-2x^3}\, \mathrm{d}x$$

Using that:

$$\int f' \cdot e^f dx = e^f + c \tag{19}$$

$$\int_{-\infty}^{\infty} x^2 e^{-2x^3} dx = \lim_{\substack{A \to -\infty \\ B \to \infty}} \int_{\pi}^{B} x^2 e^{-2x^3} dx \tag{20}$$

knowing that:

$$\begin{aligned} f &= -2x^3 \\ f' &= -6x^2 \end{aligned} \tag{21}$$

$$= \lim_{\substack{A \to -\infty \\ B \to \infty}} -\frac{1}{6} \int_{A}^{B} -6x^2 e^{-2x^3} dx = \lim_{\substack{A \to -\infty \\ B \to \infty}} \left[-\frac{1}{6} e^{-2x^3} \right]_A^B = \tag{22}$$

$$= \lim_{\substack{A \to -\infty \\ B \to \infty}} \left[-\frac{1}{6e^{2X^3}} \right]_A^B = \lim_{\substack{A \to -\infty \\ B \to \infty}} \left[-\frac{1}{6e^{2B^3}} + \frac{1}{6e^{2A^3}} \right] = -\infty \tag{23}$$

Since the first element tends to 0 and the second element tends to minus infinity.

2.8 - Solution

The definition of the gamma function is

$$\Gamma(a) = \int_0^{\infty} x^{a-1} \mathrm{e}^{-x} \, \mathrm{d}x, a \in R_{>0}.$$

It is well known that there is no closed-form antiderivative expression. Show that, for $a > 1$,

$$\Gamma(a) = (a-1)\Gamma(a-1).$$

Using partial integration with $f'(x) = x^{a-1}$ and $g(x) = e^{-x}$:

$$\Gamma(a) = \int_0^{\infty} x^{a-1} \mathrm{e}^{-x} \, \mathrm{d}x = [-x^{a-1} \mathrm{e}^{-x}]_0^{\infty} - (a-1) * (-1) * \int_0^{\infty} x^{a-2} \mathrm{e}^{-x} \, \mathrm{d}x =$$

$$= 0 - 0 + (a-1) * \int_0^{\infty} x^{a-2} \mathrm{e}^{-x} \, \mathrm{d}x = (a-1)\Gamma(a-1)$$

3.1 - Solution

Let $X_1 \sim \mathrm{Bin}(n_1, p)$ independent of $X_2 \sim \mathrm{Bin}(n_2, p)$, under the usual restriction of the parameter range, and let $S = X_1 + X_2$. State the distribution of S and its expected value and variance.

Mean:

Assuming X_1 and X_2 are independent, $S = X_1 + X_2$ has mean $E[X_1] + E[X_2] = n_{x1}P + n_{x2}P$.

Variance:

VAR(S) $=\text{Var}(X_1)+\text{Var}(X_2)=n_{X1}*p(1-p)+n_{X2}*p(1-p)$

Distribution:

$S\sim Bin(n_1+n_2,p)$

3.2 - Solution

2. Let $X_1\sim\text{Bin}(n_1,p_1)$ independent of $X_2\sim\text{Bin}(n_2,p_2)$, under the usual restriction of the parameter range, and with $p_1\neq p_2$. Let $S=X_1+X_2$. Write a computable formula for the cdf of S.

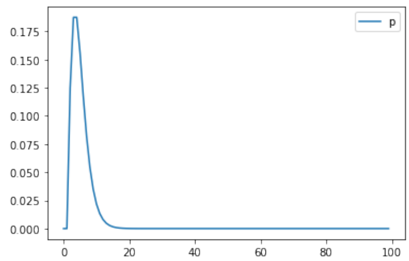
3.3 - Solution

Assume you and your spouse decide to have children until you amass 3 boys. Ignore any obvious biological constraints! Further assume p is the probability of getting a boy on each trial. Compute the expected number of children associated with this family planning strategy.

$$\sum_{k=2}^{\infty}\left(\binom{k}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{(k-2)}\frac{1}{2}\right)(k+1)=6 \tag{24}$$

Source: Python codes / 3.3 Solution

Figure 3: Distribution of the random variable describing that the xth child is the third boy



3.4 - Solution

Let $X\sim\text{Norm}\left(0,\sigma^2\right)$ Without calculation, state $\text{E}\left[X^2\right],\text{E}\left[X^3\right],\ \text{E}\left[X^4\right]$

$E\left[X^2\right]$:

$$Var(X)=E\left[X^2\right]-E[X]^2 \tag{25}$$

$$E\left[X^2\right] = Var(X) + E[X]^2,\ where\ E[X] = 0 \tag{26}$$

$$E\left[X^2\right] = Var(X) = \sigma^2 \tag{27}$$

$E\left[X^3\right]$:

By symmetry of the standard normal around zero that expected value is 0; the + and - "values" have the same "density values" so they would cancel out

$E\left[X^4\right]$:

Let $Z = X^2 \sim \chi^2$ with 1 degree of freedom. So, $E[Z] = 1$ and $Var[Z] = 2$, then:

$$\mathrm{Var}[Z] = E\left(Z^2\right) - [E[Z]]^2 = E\left[X^4\right] - \left[E\left(x^2\right)\right]^2\ where\ E\left[x^2\right] = \sigma = 1 \tag{28}$$

$$2 = E\left[X^4\right] - 1 => E\left[X^4\right] = 3 \tag{29}$$

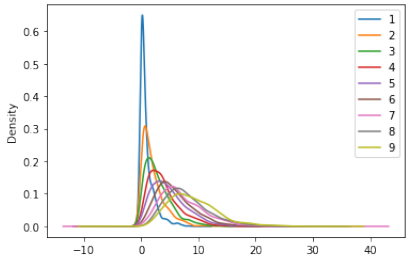
3.5 / A - Solution

χ_n^2 **:**

Randomly select several values, say 10, from a standard Normal distribution. That's a Normal distribution with a mean of 0 and standard deviation of 1. Now square each value and sum them all up. That sum is an observation from a chi-square distribution with 10 degrees of freedom. If you picked 12 values and did the same, you would have an observation from a chi-square distribution with 12 degrees of freedom.

Source: Python codes / 3.5.A Solution

Figure 4: χ_n^2 realisations with different degrees of freedom



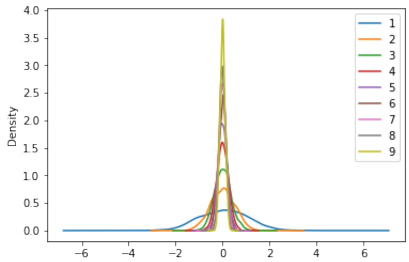
3.5 / B - Solution

Student's t:

Randomly select several values, say 10, from a standard Normal distribution. That's a Normal distribution with a mean of 0 and standard deviation of 1. Now sum up all values and divide by the root of the sample size (n). That value is an observation from a Student-t distribution with 10 degrees of freedom. If you picked 12 values and did the same, you would have an observation from a Student-t distribution with 12 degrees of freedom.

Source: Python codes / 3.5.B Solution

Figure 5: Student t realisations with different degrees of freedom



3.6 - Solution

Recall that a probability measure is a set function that assigns a real number $\Pr(A)$ to each event $A \in \mathcal{A}$, and has 3 requirements. State them, and describe what is meant here by $\mathcal{A}$.

(i) Non negativity $P : \mathcal{A} \to [0, 1]$

(ii) $P(\Omega) = 1$ and $P(\emptyset) = 1$

(iii) Sigma additivity or infinite additivity: $P\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P\left(A_n\right)$

3.7 - Solution

Given the characteristic function of a continuous univariate random variable, explain in words how you can compute the pdf.

One can transform characteristic funtion to pdf function useing Furier-transform.

CF- Characteristic Function

PDF - Probability Density Function

CDF - Cumulative Density Function

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t) dt$$

3.8 - Solution

In words, explain the (weak) law of large numbers, and the central limit theorem, and how they differ.

Central Limit Theorem (CLT) - states that for multiple samples taken from a population (with known mean and variance), if the sample size is large, then the distribution of the sample mean, or sum, will converge to a normal distribution even though the random variable x (individual data points within a sample) may be non-normal. This proves to be a key concept in probability theory as it implies that probabilistic and statistical methods that work for normal distributions can be applicable to many problems involving other types of distributions. It usually gives the below conditions

1. Sample means always follow normal distribution irrespective of distribution of individual data in population 2. Mean of sample means tends to population mean as the number of samples tend to infinity 3. Variance of sample means is 'n' times less than the variance of population, where 'n' is size of sample e.g. Consider the roll of 2 dice. If this is done multiple times and the average or the sum of the rolls is plotted, then this plot will converge to a normal distribution

Law of Large Numbers - states that as sample size grows, the sample mean gets closer to the population mean irrespective whether the data set is normal or non-normal e.g. consider the roll of a single dice. If you roll the dice sufficiently large number of times, the average would tend to be close to 3.5.

3.9 - Solution

State 5 common univariate discrete distributions. (You do not need to write out the mass functions.)

(i) **Indicator distribution**

- Value set: $\{0, 1\}$
- Distribution: $\begin{matrix} 1-p & \text{if } k=0 \\ p & \text{if } k=1 \end{matrix}$
- Expected value: p
- Variance: p(1-p)

(ii) **Binomial distribution**

- Value set: $\{0, .., n\}$
- Distribution: $P(X = k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$ for $k = 0, 1, 2..., n$
- Expected value: n*p
- Variance: n*p(1-p)

(iii) Hypergeometric distribution

- Value set: $\{0, .., n\}$
- Distribution: $P(x) = \dfrac{\begin{pmatrix} k \\ x \end{pmatrix}\begin{pmatrix} N-k \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$
- Expected value: $\mu = E(X) = \frac{nk}{N}$
- Variance: $\sigma^2 = V(X) = \left(\frac{N-n}{N-1}\right)(n)\left(\frac{k}{N}\right)\left(\frac{N-k}{N}\right)$

(iv) Poission distribution

- Value set: $\{0, 1, 2, 3, ..\}$
- Distribution: $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ Where: -> X is a random variable following a Poisson distribution

-> k is the number of times an event occurs

-> $P(X = k)$ is the probability that an event will occur k times

-> e is Euler's constant (approximately 2.718)

-> λ is the average number of times an event occurs

- Expected value: λ

- Variance: λ

(v) Geometric distribution

- Value set: $\{1, 2, 3, ..\}$
- Distribution: $p_X(x) = \begin{cases} (1-p)^x p & \text{if } x \in R_X \\ 0 & \text{if } x \notin R_X \end{cases}$
- Expected value: $\mathrm{E}[X] = \frac{1-p}{p}$
- Variance: $\mathrm{Var}[X] = \frac{1-p}{p^2}$

(vi) negative binomial distribution

- Value set: $\{1, 2, 3, ..\}$
- Distribution: $\begin{pmatrix} k-1 \\ r-1 \end{pmatrix} (1-p)^{k-r} p^r$
- Expected value: $\mathrm{E}[X] = \frac{r}{p}$
- Variance: $r\frac{1-p}{p^2}$

3.10 - Solution

State a (common) distribution whose expectation does not exist.

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$$

3.11 - Solution

State the pdf and cdf of an exponential random variable with rate λ . Given a set of IID exponential random variables, how is their sum distributed?

PDF:

$$F(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ 1 - e^{-\lambda * x}, & \text{if } 0 < x \end{cases}$$

CDF:

$$f(x) = \begin{cases} 0, & \text{if} \quad x \leq 0 \\ \lambda e^{-\lambda x}, & \text{if} \quad 0 < x \end{cases}$$

The sum of n independent exponential random variables is a Gamma random variable $\sim \Gamma(n, \lambda)$.

3.12 - Solution

(one question on linear algebra) For a real symmetric matrix, state relationships between the eigenvalues, the trace, and the determinant of the matrix.

Theorem: If A is an $n \times n$ matrix, then the sum of the n eigenvalues of A is the trace of A and the product of the n eigenvalues is the determinant of A. Proof: Write $A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$. Also let the n eigenvalues of A be $\lambda_1, \dots, \lambda_n$. Finally, denote the characteristic polynomial of A by $p(\lambda) = |\lambda I - A| = \lambda^n + c_{n-1}\lambda^{n-1} + \dots + c_1\lambda + c_0$. Note that since the eigenvalues of A are the zeros of $p(\lambda)$, this implies that $p(\lambda)$ can be factorised as $p(\lambda) = (\lambda - \lambda_1) \dots (\lambda - \lambda_n)$.

Consider the constant term of $p(\lambda), c_0$. The constant term of $p(\lambda)$ is given by $p(0)$, which can be calculated in two ways:

Firstly, $p(0) = (0 - \lambda_1) \dots (0 - \lambda_n) = (-1)^n \lambda_1 \dots \lambda_n$. Secondly, $p(0) = |0I - A| = |-A| = (-1)^n |A|$. Therefore $c_0 = (-1)^n \lambda_1 \dots \lambda_n = (-1)^n |A|$, and so $\lambda_1 \dots \lambda_n = |A|$. That is, the product of the n eigenvalues of A is the determinant of A. Consider the coefficient of λ^{n-1}, c_{n-1}. This is also calculated in two ways. Firstly, it can be calculated by expanding $p(\lambda) = (\lambda - \lambda_1) \dots (\lambda - \lambda_n)$. In order to get the λ^{n-1} term, the λ must be chosen from $n - 1$ of the factors, and the constant from the other.

Hence, the λ^{n-1} term will be $-\lambda_1\lambda^{n-1} - \cdots - \lambda\lambda^{n-1} = -(\lambda_1 + \cdots + \lambda_n)\lambda^{n-1}$. Thus $c_{n-1} = -(\lambda_1 + \cdots + \lambda_n)$. Secondly, this coefficient can be calculated by expanding $|\lambda I - A|$:

$$|\lambda I - A| = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & \lambda - a_{nn} \end{vmatrix}$$

One way of calculating determinants is to multiply the elements in positions $1j_1, 2j_2, \ldots, nj_n$, for each possible permutation $j_1 \ldots j_n$ of $1 \ldots n$. If the permutation is odd, then the product is also multiplied by -1. Then all of these n ! products are added together to produce the determinant. One of these products is $(\lambda - a_{11}) \ldots (\lambda - a_{nn})$. Every other possible product can contain at most $n-2$ elements on the diagonal of the matrix, and so will contain at most $n-2\lambda$'s. Hence, when all of these other products are expanded, they will produce a polynomial in λ of degree at most $n-2$. Denote this polynomial by $q(\lambda)$.

Hence, $p(\lambda) = (\lambda - a_{11}) \ldots (\lambda - a_{nn}) + q(\lambda)$. Since $q(\lambda)$ has degree at most $n-2$, it has no λ^{n-1} term, and so the λ^{n-1} term of $p(\lambda)$ must be the λ^{n-1} term from $(\lambda - a_{11}) \ldots (\lambda - a_{nn})$. However, the argument above for $(\lambda - \lambda_1) \ldots (\lambda - \lambda_n)$ shows that this term must be $-(a_{11} + \cdots + a_{nn})\lambda^{n-1}$. Therefore $c_{n-1} = -(\lambda_1 + \cdots + \lambda_n) = -(a_{11} + \cdots + a_{nn})$, and so $\lambda_1 + \cdots + \lambda_n = a_{11} + \cdots + a_{nn}$. That is, the sum of the n eigenvalues of A is the trace of A.

STOCHASTIC

1. The conditional expectation will be a random variable that takes only two distinct values: for every $a \in A$, it will take the value of the average of the original random variable's values on A weighted by the respective probabilities, and for every $a \in \Omega - A$, it will take the value of the average of the original random variable's values on $\Omega - A$ weighted by the respective probabilities

2. It gives the best estimate in terms of mean squared error. If we consider all

the possible constructions of a random variable based only on the information we have, and consider the difference between this and the original random variable, the conditional expectation will have the minimal difference among all those possibilities.

3.

$$E\left[S_n|\sigma(X_1,...,X_{n-1})\right] = E\left[\sum_{i=1}^{n} X_i|\sigma(X_1,...,X_{n-1})\right] =$$

$$= E\left[\sum_{i=1}^{n-1} X_i + X_n|\sigma(X_1,...,X_{n-1})\right] = \sum_{i=1}^{n-1} X_i + E\left[X_n|\sigma(X_1,...,X_{n-1})\right] =$$

$$= E\left[\sum_{i=1}^{n-1} X_i + X_n|\sigma(X_1,...,X_{n-1})\right] = \sum_{i=1}^{n-1} X_i + E\left[X_n\right] = \sum_{i=1}^{n-1} X_i = S_{n-1}$$

using that $\sum_{i=1}^{n-1} X_i$ is $\sigma(X_1,...,X_{n-1})$-measurable and X_n is independent of $\sigma(X_1,...,X_{n-1})$.

This is reasonable because each step has 0 expected value, so we expect that the best estimate of where we will be 1 step after is where are we at the current step.

4.

$$E\left[S_n|\sigma(X_1,...,X_{n-1})\right] = E\left[\sum_{i=1}^{n} X_i|\sigma(X_1,...,X_{n-1})\right] =$$

$$= E\left[\sum_{i=1}^{n-1} X_i + X_n|\sigma(X_1,...,X_{n-1})\right] = \sum_{i=1}^{n-1} X_i + E\left[X_n|\sigma(X_1,...,X_{n-1})\right] =$$

$$= E\left[\sum_{i=1}^{n-1} X_i + X_n|\sigma(X_1,...,X_{n-1})\right] = \sum_{i=1}^{n-1} X_i + E\left[X_n\right] = \sum_{i=1}^{n-1} X_i + \frac{1}{3} = S_{n-1} + \frac{1}{3}$$

5. $E[\tau] = 2 * 5 = 10$ because

6. The number of paths going from A to B which touch or cross the t axis is equal to the number of paths going from A' to B, where A' is the reflection of A with respect to the t axis.

7. A discrete time martingale is a sequence $X_1, ..., X_n, ...$ of random variables on a common probability space which, for all n satisfy

$$E\left[X_{n+1}|X_1,...,X_n\right] = X_n$$

Yes, the S_n from Question 3 is a martingale.

8. Strong law of large numbers:

$$P\left(lim_{n\to\infty}\frac{\sum_{i=1}^{n} X_i}{n} = \mu\right) = 1$$

Central Limit Theorem:

$$\frac{\frac{\sum_{i=1}^{n} X_i}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1)$$

in distribution.

The strong law of large numbers says that the average of the sample converges to the expected value with probability 1. The Central Limit Theorem says that a properly demeaned and scaled version of the sample approximates a standard normal distribution as the sample size grows big enough.

9. If we sample from a distribution which has no finite expected value, for example the Cauchy-distribution, then the law of large numbers will fail. In the case of the Cauchy distribution, the average will not converge to anything, it will "jump around forever".

10. Two random variables are uncorrelated when there is no linear dependence between them, and independent when there are no dependency of any kind between them. The strong law of large numebrs holds in the case also when the variables are pairwise uncorrelated.

11. We can calculate the itnegral with the Monte Carlo Method the following way: Sample n points uniformly from $[0,1]^d$, let those be $x_1, x_2, ..., x_n$. Then we can approximate the value of the integral with $\frac{x_1+x_2+...+x_n}{n}$, which is converging to the integral value by the law of large numbers with probability one (if the integral exists and is finite). Pseudocode:

    I $= 0$

    for i from 1 to n:

    $x[i] = random.uniform([0,1]^d)$

    for i from 1 to n:

I += f(x[i])

I = I / n

If we use the notation $\sigma^2(f) = \int_{[0,1]^d} f^2(x)dx - \left(\int_{[0,1]^d} f(x)dx\right)^2$, then by additivity of variance it turn out that the expected error for the Monte Carlo integral is $\frac{\sigma(f)}{\sqrt{n}}$.

These are all probabilistic statements.

12. No, the convergence rate only depends on the variance of f. This is useful because most deterministic methods converge exponentially slower as d increases but for Monte Carlo it is at the same rate, which makes it useful to calculate integrals in large dimensions.

Python Codes

3.3

```
import math
import pandas as pd

def binom(n, k):
    return math.factorial(n) // math.factorial(k) // math.factorial(n - k)

max=100
sol=[0]*(max)
for n in range(2,max):
    sol[n]=binom(n, 2)*(1/2)**2*(1/2)**(n-2)*(1/2)
sol=pd.DataFrame(sol)
sol.columns=["p"]
sol.plot()

# calculate probability
sol["weg"]=""
for i in range(0,len(sol)):
    sol["weg"][i]=sol["p"].iloc[i]*(i+1)
sum(sol['weg'])
```

3.5 / A

```
import seaborn as sns
import numpy as np
import statistics

def square(list):
    return map(lambda x: x ** 2, list)
```

```
freedom_range=[*range(1,10,1)]
norms=[0]*n
chi=[0]*1000
df = pd.DataFrame() #pd.DataFrame(columns=[freedom_range])
for i in freedom_range:
    for j in range(0,1000):
        n=i
        norms=np.random.normal(0, 1, n)
        chi[j]=sum(square(norms))
    df[str(i)]=chi
df.plot(kind='density')
```

3.5 / B

```
import seaborn as sns
import numpy as np
import statistics

def square(list):
    return map(lambda x: x ** 2, list)

freedom_range=[*range(1,10,1)]
norms=[0]*n
ts=[0]*1000
df = pd.DataFrame() #pd.DataFrame(columns=[freedom_range])
for i in freedom_range:
    for j in range(0,1000):
        n=i
        norms=np.random.normal(0, 1, n)
        ts[j]=statistics.mean(norms)/math.sqrt(n)
    df[str(i)]=ts
df.plot(kind='density')
```