10-19-2015

04. Random Variables: Concepts

Gerhard Müller

University of Rhode Island, gmuller@uri.edu

Abstract

Part four of course materials for Nonequilibrium Statistical Physics (Physics 626), taught by Gerhard Müller at the University of Rhode Island. Entries listed in the table of contents, but not shown in the document, exist only in handwritten form. Documents will be updated periodically as more entries become presentable.

Updated with version 2 on 5/3/2016.

Contents of this Document [ntc4]

Probability Distribution [nln46]

Experiment represented by events in a sample space: $S = \{A_1, A_2, \ldots\}$.

Measurements represented by stochastic variable: $X = \{x_1, x_2, \ldots\}$.

Maximum amount of information experimentally obtainable is contained in the probability distribution:

$$P_X(x_i) \geq 0, \quad \sum_i P_X(x_i) = 1.$$

Partial information is contained in moments,

$$\langle X^n \rangle = \sum_i x_i^n P_X(x_i), \quad n = 1, 2, \ldots,$$

or cumulants (as defined in [nln47]),

- $\langle\langle X \rangle\rangle = \langle X \rangle$ (mean value)
- $\langle\langle X^2 \rangle\rangle = \langle X^2 \rangle - \langle X \rangle^2$ (variance)
- $\langle\langle X^3 \rangle\rangle = \langle X^3 \rangle - 3\langle X \rangle\langle X^2 \rangle + 2\langle X \rangle^3$

The variance is the square of the standard deviation: $\langle\langle X^2 \rangle\rangle = \sigma_X^2$.

For continuous stochastic variables we have

$$P_X(x) \geq 0, \quad \int dx P_X(x) = 1, \quad \langle X^n \rangle = \int dx\, x^n P(x).$$

In the literature $P_X(x)$ is often named 'probability density' and the term 'distribution' is used for

$$F_X(x) = \sum_{x_i < x} P_X(x_i) \quad \text{or} \quad F_X(x) = \int_{-\infty}^{x} dx' P_X(x')$$

in a cumulative sense.

Characteristic Function [nln47]

Fourier transform: $\Phi_X(k) \doteq \left\langle e^{ikx} \right\rangle = \int_{-\infty}^{+\infty} dx\, e^{ikx} P_X(x).$

Attributes: $\Phi_X(0) = 1, \quad |\Phi_X(k)| \leq 1.$

Moment generating function:

$$\Phi_X(k) = \int_{-\infty}^{+\infty} dx\, P_X(x) \left[\sum_{n=0}^{\infty} \frac{(ik)^n}{n!} x^n \right] = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \langle X^n \rangle$$

$$\Rightarrow \ \langle X^n \rangle \doteq \int_{-\infty}^{+\infty} dx\, x^n P_X(x) = (-i)^n \frac{d^n}{dk^n} \Phi_X(k) \bigg|_{k=0}.$$

Cumulant generating function:

$$\ln \Phi_X(k) \doteq \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \langle\langle X^n \rangle\rangle \quad \Rightarrow \ \langle\langle X^n \rangle\rangle = (-i)^n \frac{d^n}{dk^n} \ln \Phi_X(k) \bigg|_{k=0}.$$

Cumulants in terms of moments (with $\Delta X \doteq X - \langle X \rangle$): [nex126]

- $\langle\langle X \rangle\rangle = \langle X \rangle$
- $\langle\langle X^2 \rangle\rangle = \langle X^2 \rangle - \langle X \rangle^2 = \langle (\Delta X)^2 \rangle$
- $\langle\langle X^3 \rangle\rangle = \langle (\Delta X)^3 \rangle$
- $\langle\langle X^4 \rangle\rangle = \langle (\Delta X)^4 \rangle - 3\langle (\Delta X)^2 \rangle^2$

Theorem of Marcienkiewicz:
$\ln \Phi_X(k)$ can only be a polynomial if the degree is $n \leq 2$.

- $n = 1$: $\ln \Phi_X(k) = ika \quad \Rightarrow \ P_X(x) = \delta(x - a)$
- $n = 2$: $\ln \Phi_X(k) = ika - \frac{1}{2}bk^2 \quad \Rightarrow \ P_X(x) = \frac{1}{\sqrt{2\pi b}} \exp\left(-\frac{(x-a)^2}{2b} \right)$

Consequence: any probability distribution has either one, two, or infinitely many non-vanishing cumulants.

[nex126] Cumulants expressed in terms of moments

The characteristic function $\Phi_X(k)$ of a probability distribution $P_X(x)$, obtained via Fourier transform as described in [nln47], can be used to generate the moments $\langle X^n \rangle$ and the cumulants $\langle\langle X^n \rangle\rangle$ via the expansions

$$\Phi_X(k) = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \langle X^n \rangle, \qquad \ln \Phi_X(k) = \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \langle\langle X^n \rangle\rangle.$$

Use these relations to express the first four cumulants in terms of the first four moments. The results are stated in [nln47]. Describe your work in some detail.

Solution:

Generating function [nln48]

The generating function $G_X(z)$ is a representation of the characteristic function $\Phi_X(k)$ that is most commonly used, along with factorial moments and factorial cumulants, if the stochastic variable X is integer valued.

Definition: $G_X(z) \doteq \langle z^x \rangle$ with $|z| = 1$.

Application to continuous and discrete (integer-valued) stochastic variables:

$$G_X(z) = \int dx\, z^x P_X(x), \qquad G_X(z) = \sum_n z^n P_X(n).$$

Definition of factorial moments:

$$\langle X^m \rangle_f \doteq \langle X(X-1)\cdots(X-m+1)\rangle, \quad m \geq 1; \quad \langle X^0 \rangle_f \doteq 0.$$

Function generating factorial moments:

$$G_X(z) = \sum_{m=0}^{\infty} \frac{(z-1)^m}{m!} \langle X^m \rangle_f, \quad \langle X^m \rangle_f = \frac{d^m}{dz^m} G_X(z)\bigg|_{z=1}.$$

Function generating factorial cumulants:

$$\ln G_X(z) = \sum_{m=1}^{\infty} \frac{(z-1)^m}{m!} \langle\langle X^m \rangle\rangle_f, \quad \langle\langle X^m \rangle\rangle_f = \frac{d^m}{dz^m} \ln G_X(z)\bigg|_{z=1}.$$

Applications:

- ▷ Moments and cumulants of the Poisson distribution [nex16]
- ▷ Pascal distribution [nex22]
- ▷ Reconstructing probability distributions [nex14]

Multivariate Distributions [nln7]

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector variable with n components.

Joint probability distribution: $P(x_1, \ldots, x_n)$.

Marginal probability distribution:

$$P(x_1, \ldots, x_m) = \int dx_{m+1} \cdots dx_n \, P(x_1, \ldots, x_n).$$

Conditional probability distribution: $P(x_1, \ldots, x_m | x_{m+1}, \ldots, x_n)$.

$$P(x_1, \ldots, x_n) = P(x_1, \ldots, x_m | x_{m+1}, \ldots, x_n) P(x_{m+1}, \ldots, x_n).$$

Moments: $\langle X_1^{m_1} \cdots X_n^{m_n} \rangle = \int dx_1 \cdots dx_n \, x_1^{m_1} \cdots x_n^{m_n} P(x_1, \ldots, x_n)$.

Characteristic function: $\Phi(\mathbf{k}) = \langle e^{i\mathbf{k}\cdot\mathbf{X}} \rangle$.

Moment expansion: $\displaystyle \Phi(\mathbf{k}) = \sum_0^\infty \frac{(ik_1)^{m_1} \cdots (ik_n)^{m_n}}{m_1! \ldots m_n!} \langle X_1^{m_1} \cdots X_n^{m_n} \rangle$.

Cumulant expansion: $\displaystyle \ln \Phi(\mathbf{k}) = {\sum_0^\infty}' \frac{(ik_1)^{m_1} \cdots (ik_n)^{m_n}}{m_1! \ldots m_n!} \langle\langle X_1^{m_1} \cdots X_n^{m_n} \rangle\rangle$.

(prime indicates absence of term with $m_1 = \cdots = m_n = 0$).

Covariance matrix: $\langle\langle X_i X_j \rangle\rangle = \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle$.
($i = j$: variances, $i \neq j$: covariances).

Correlations: $\displaystyle C(X_i, X_j) = \frac{\langle\langle X_i X_j \rangle\rangle}{\sqrt{\langle\langle X_i \rangle\rangle \langle\langle X_j \rangle\rangle}}$.

Statistical independence of X_1, X_2: $P(x_1, x_2) = P_1(x_1) P_2(x_2)$.

Equivalent criteria for statistical independence:

- all moments factorize: $\langle X_1^{m_1} X_2^{m_2} \rangle = \langle X_1^{m_1} \rangle \langle X_2^{m_2} \rangle$;
- characteristic function factorizes: $\Phi(k_1, k_2) = \Phi_1(k_1) \Phi_2(k_2)$;
- all cumulants $\langle\langle X_1^{m_1} X_2^{m_2} \rangle\rangle$ with $m_1 m_2 \neq 0$ vanish.

If $\langle\langle X_1 X_2 \rangle\rangle = 0$ then X_1, X_2 are called *uncorrelated*.
This property does not imply *statistical independence*.

Transformation of Random Variables [nln49]

Consider two random variables X and Y that are functionally related:

$$Y = F(X) \quad \text{or} \quad X = G(Y).$$

If the probability distribution for X is known then the probability distribution for Y is determined as follows:

$$P_Y(y)\Delta y = \int_{y<f(x)<y+\Delta y} dx P_X(x)$$

$$\Rightarrow \; P_Y(y) = \int dx P_X(x)\delta\big(y - f(x)\big) = P_X\big(g(y)\big)\,|g'(y)|\,.$$

Consider two random variables X_1, X_2 with a joint probability distribution

$$P_{12}(x_1, x_2).$$

The probability distribution of the random variable $Y = X_1 + X_2$ is then determined as

$$P_Y(y) = \int dx_1 \int dx_2\, P_{12}(x_1, x_2)\delta(y - x_1 - x_2) = \int dx_1 P_{12}(x_1, y - x_1),$$

and the probability distribution of the random variable $Z = X_1 X_2$ as

$$P_Z(z) = \int dx_1 \int dx_2\, P_{12}(x_1, x_2)\delta(z - x_1 x_2) = \int \frac{dx_1}{|x_1|} P_{12}(x_1, z/x_1).$$

If the two random variables X_1, X_2 are statistically independent we can substitute $P_{12}(x_1, x_2) = P_1(x_1)P_2(x_2)$ in the above integrals.

Applications:

▷ Transformation of statistical uncertainty [nex24]

▷ Chebyshev inequality [nex6]

▷ Robust probability distributions [nex19]

▷ Statistically independent or merely uncorrelated? [nex23]

▷ Sum and product of uniform distributions [nex96]

▷ Exponential integral distribution [nex79]

▷ Generating exponential and Lorentzian random numbers [nex80]

▷ From Gaussian to exponential distribution [nex8]

▷ Transforming a pair of random variables [nex78]

[nex127] Sums of independent exponentials

Consider n independent random variable $X_1, \ldots, X_n$ with range $x_i \geq 0$ and identical exponential distributions,

$$P_1(x_i) = \frac{1}{\xi}\, e^{-x_i/\xi}, \quad i = 1, \ldots, n.$$

Use the transformation relation from [nln49],

$$P_2(x) = \int dx_1 \int dx_2\, P_1(x_1)P_1(x_2)\delta(x - x_1 - x_2) = \int dx_1 P_1(x_1)P_1(x - x_1),$$

inductively to calculate the probability distribution $P_n(x)$, $n \geq 2$ of the stochastic variable

$$X = X_1 + \cdots + X_n.$$

Find the mean value $\langle X \rangle$, the variance $\langle\langle X^2 \rangle\rangle$, and the value x_p where $P_n(x)$ has its peak value.

Solution:

[nex24] Transformation of statistical uncertainty.

From a given stochastic variable X with probability distribution $P_X(x)$ we can calculate the probability distribution of the stochastic variable $Y = f(X)$ via the relation

$$P_Y(y) = \int dx\, P_X(x)\delta\left(y - f(x)\right).$$

Show by systematic expansion that if $P_X(x)$ is sufficiently narrow and $f(x)$ sufficiently smooth, then the mean values and the standard deviations of the two stochastic variables are related to each other as follows:

$$\langle Y\rangle = f(\langle X\rangle), \qquad \sigma_Y = |f'(\langle X\rangle)|\sigma_X.$$

Solution:

[nex6] Chebyshev's inequality

Chebyshev's inequality is a rigorous relation between the standard deviation $\sigma_X = \sqrt{\langle X^2 \rangle - \langle X \rangle^2}$ of the random variable X and the probability of deviations from the mean value $\langle X \rangle$ greater than a given magnitude a.

$$P[(x - \langle X \rangle)^2 > a^2] \leq \left(\frac{\sigma_X}{a}\right)^2$$

Prove Chebyshev's inequality starting from the following relation, commonly used for the transformation of stochastic variables (as in [nln49]):

$$P_Y(y) = \int dx\, \delta(y - f(x)) P_X(x) \text{ with } f(x) = (x - \langle X \rangle)^2.$$

Solution:

[nex7] Law of large numbers

Let $X_1, \ldots, X_N$ be N statistically independent random variables described by the same probability distribution $P_X(x)$ with mean value $\langle X \rangle$ and standard deviation $\sigma_X = \sqrt{\langle X^2 \rangle - \langle X \rangle^2}$. These random variables might represent, for example, a series of measurements under the same (controllable) conditions. The law of large numbers states that the uncertainty (as measured by the standard deviation) of the stochastic variable $Y = (X_1 + \cdots + X_N)/N$ is

$$\sigma_Y = \frac{\sigma_X}{\sqrt{N}}.$$

Prove this result.

Solution:

Binomial, Poisson, and Gaussian Distributions [nln8]

Consider a set of N independent experiments, each having two possible outcomes occurring with given probabilities.

events	$A + B = S$
probabilities	$p + q = 1$
random variables	$n + m = N$

Binomial distribution:

$$P_N(n) = \frac{N!}{n!(N-n)!}\, p^n (1-p)^{N-n}.$$

Mean value: $\langle n \rangle = Np$.

Variance: $\langle\langle n^2 \rangle\rangle = Npq$. [nex15]]

In the following we consider two different asymptotic distributions in the limit $N \to \infty$.

Poisson distribution:

Limit #1: $N \to \infty$, $p \to 0$ such that $Np = \langle n \rangle = a$ stays finite [nex15].

$$P(n) = \frac{a^n}{n!}\, e^{-a}.$$

Cumulants: $\langle\langle n^m \rangle\rangle = a$.

Factorial cumulants: $\langle\langle n^m \rangle\rangle_f = a\delta_{m,1}$. [nex16]

Single parameter: $\langle n \rangle = \langle\langle n^2 \rangle\rangle = a$.

Gaussian distribution:

Limit #2: $N \gg 1$, $p > 0$ with $Np \gg \sqrt{Npq}$.

$$P_N(n) = \frac{1}{\sqrt{2\pi\langle\langle n^2 \rangle\rangle}} \exp\left(-\frac{(n - \langle n \rangle)^2}{2\langle\langle n^2 \rangle\rangle} \right).$$

Derivation: DeMoivre-Laplace limit theorem [nex21].

Two parameters: $\langle n \rangle = Np$, $\langle\langle n^2 \rangle\rangle = Npq$.

Special case of central limit theorem [nln9].

[nex15] Binomial to Poisson distribution

Consider the binomial distribution for two events A, B that occur with probabilities $P(A) \equiv p$, $P(B) = 1 - p \equiv q$, respectively:

$$P_N(n) = \frac{N!}{n!(N-n)!}\, p^n q^{N-n},$$

where N is the number of (independent) experiments performed, and n is the stochastic variable that counts the number of realizations of event A.
(a) Find the mean value $\langle n\rangle$ and the variance $\langle\langle n^2\rangle\rangle$ of the stochastic variable n.
(b) Show that for $N \to \infty$, $p \to 0$ with $Np \to a > 0$, the binomial distribution turns into the Poisson distribution

$$P_\infty(n) = \frac{a^n}{n!}\, e^{-a}.$$

Solution:

[nex21] De Moivre–Laplace limit theorem.

Show that for large Np and large Npq the binomial distribution turns into the Gaussian distribution with the same mean value $\langle n\rangle = Np$ and variance $\langle\langle n^2\rangle\rangle = Npq$:

$$P_N(n) = \frac{N!}{n!(N-n)!}\, p^n q^{N-n} \quad \longrightarrow \quad P_N(n) \simeq \frac{1}{\sqrt{2\pi\langle\langle n^2\rangle\rangle}} \exp\left(-\frac{(n-\langle n\rangle)^2}{2\langle\langle n^2\rangle\rangle}\right).$$

Solution:

Central Limit Theorem [nln9]

The central limit theorem is a major extension of the law of large numbers. It explains the unique role of the Gaussian distribution in statistical physics.

Given are a large number of statistically independent random variables $X_i, i = 1, \ldots, N$ with equal probability distributions $P_X(x_i)$. The only restriction on the shape of $P_X(x_i)$ is that the moments $\langle X_i^n \rangle = \langle X^n \rangle$ are finite for all n.

Goal: Find the probability distribution $P_Y(y)$ for the random variable $Y = (X_1 - \langle X \rangle + \cdots + X_N - \langle X \rangle)/N$.

$$P_Y(y) = \int dx_1 \, P_X(x_1) \cdots \int dx_N \, P_X(x_N) \delta\left(y - \frac{1}{N} \sum_{i=1}^{N} [x_i - \langle X \rangle] \right).$$

Characteristic function:

$$\Phi_Y(k) \equiv \int dy \, e^{iky} P_Y(y), \quad P_Y(y) = \frac{1}{2\pi} \int dk \, e^{-iky} \Phi_Y(k).$$

$$\begin{aligned} \Rightarrow \ \Phi_Y(k) &= \int dx_1 \, P_X(x_1) \cdots \int dx_N \, P_X(x_N) \exp\left(i \frac{k}{N} \sum_{i=1}^{N} [x_i - \langle X \rangle] \right) \\ &= \left[\bar{\Phi}\left(k/N \right) \right]^N, \end{aligned}$$

$$\begin{aligned} \bar{\Phi}\left(\frac{k}{N} \right) &= \int dx \, e^{i(k/N)(x - \langle X \rangle)} P_X(x) = \exp\left(-\frac{1}{2} \left(\frac{k}{N} \right)^2 \langle\langle X^2 \rangle\rangle + \cdots \right) \\ &= 1 - \frac{1}{2} \left(\frac{k}{N} \right)^2 \langle\langle X^2 \rangle\rangle + \mathrm{O}\left(\frac{k^3}{N^3} \right), \end{aligned}$$

where we have performed a cumulant expansion to leading order.

$$\Rightarrow \ \Phi_Y(y) = \left[1 - \frac{k^2 \langle\langle X^2 \rangle\rangle}{2N^2} + \mathrm{O}\left(\frac{k^3}{N^3} \right) \right]^N \stackrel{N\to\infty}{\longrightarrow} \exp\left(-\frac{k^2 \langle\langle X^2 \rangle\rangle}{2N} \right).$$

where we have used $\lim_{N\to\infty}(1 + z/N)^N = e^z$.

$$\Rightarrow \ P_Y(y) = \sqrt{\frac{N}{2\pi \langle\langle X^2 \rangle\rangle}} \exp\left(-\frac{Ny^2}{2\langle\langle X^2 \rangle\rangle} \right) = \frac{1}{\sqrt{2\pi \langle\langle Y^2 \rangle\rangle}} e^{-y^2/2\langle\langle Y^2 \rangle\rangle}$$

with variance $\langle\langle Y^2 \rangle\rangle = \langle\langle X^2 \rangle\rangle / N$

Note that regardless of the form of $P_X(x)$, the average of a large number of (independent) measurements of X will be a Gaussian with standard deviation $\sigma_Y = \sigma_X/\sqrt{N}$.

[nex19] Robust probability distributions

Consider two independent stochastic variables X_1 and X_2, each specified by the same probability distribution $P_X(x)$. Show that if $P_X(x)$ is either a Gaussian, a Lorentzian, or a Poisson distribution,

$$\text{(i)}\;\; P_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-x^2/2\sigma^2}, \qquad \text{(ii)}\;\; P_X(x) = \frac{1}{\pi}\frac{a}{x^2+a^2}, \qquad \text{(iii)}\;\; P_X(x=n) = \frac{a^n}{n!}\, e^{-a}.$$

then the probability distribution $P_Y(y)$ of the stochastic variable $Y = X_1 + X_2$ is also a Gaussian, a Lorentzian, or a Poisson distribution, respectively. What property of the characteristic function $\Phi_X(k)$ guarantees the robustness of $P_X(x)$?

Solution:

[nex81] Stable probability distributions

Consider N independent random variables $X_1, \ldots, X_N$, each having the same probability distribution $P_X(x)$. If the probability distribution of the random variable $Y_N = X_1 + \cdots + X_N$ can be written in the form $P_Y(y) = P_X(y/c_N + \gamma_N)/c_N$, then $P_X(x)$ is *stable*. The multiplicative constant must be of the form $c_N = N^{1/\alpha}$, where α is the *index* of the stable distribution. $P_X(x)$ is *strictly stable* if $\gamma_N = 0$.
Use the results of [nex19] to determine the indices α of the Gaussian and Lorentzian distributions, both of which are both strictly stable. Show that the Poisson distribution is not stable in the technical sense used here.

Solution:

Exponential distribution [nln10]

Busses arrive randomly at a bus station.
The average interval between successive bus arrivals is τ.

$f(t)dt$: probability that the interval is between t and $t+dt$.

$P_0(t) = \int_t^\infty dt' f(t')$: probability that the interval is larger than t.

Relation: $f(t) = -\dfrac{dP_0}{dt}$.

Normalizations: $P_0(0) = 1, \quad \int_0^\infty dt\, f(t) = 1$.

Mean value: $\langle t \rangle \equiv \int_0^\infty dt\, t f(t) = \tau$.

Start the clock when a bus has arrived and consider the events A and B.

Event A: the next bus has not arrived by time t.
Event B: a bus arrives between times t and $t+dt$.

Assumptions:

1. $P(AB) = P(A)P(B)$ (statistical independence).
2. $P(B) = cdt$ with c to be determined.

Consequence: $P_0(t+dt) = P(A\bar{B}) = P(A)P(\bar{B}) = P_0(t)[1-cdt]$.

$$\Rightarrow \frac{d}{dt}P_0(t) = -cP_0(t) \;\Rightarrow\; P_0(t) = e^{-ct} \;\Rightarrow\; f(t) = ce^{-ct}.$$

Adjust mean value: $\langle t \rangle = \tau \;\Rightarrow\; c = 1/\tau$.

Exponential distribution: $P_0(t) = e^{-t/\tau}, \quad f(t) = \dfrac{1}{\tau}e^{-t/\tau}$.

Find the probability $P_n(t)$ that n busses arrive before time t.

First consider the probabilities $f(t')dt'$ and $P_0(t-t')$ of the two statistically independent events that the first bus arrives between t' and $t'+dt'$ and that no futher bus arrives until time t.

Probability that exactly one bus arrives until time t:

$$P_1(t) = \int_0^t dt' f(t') P_0(t-t') = \frac{t}{\tau}e^{-t/\tau}.$$

Then calculate $P_n(t)$ by induction.

Poisson distribution: $P_n(t) = \int_0^t dt' f(t') P_{n-1}(t-t') = \dfrac{(t/\tau)^n}{n!}e^{-t/\tau}$.

Waiting Time Problem [nln11]

Busses arrive more or less randomly at a bus station.
Given is the probability distribution $f(t)$ for intervals between bus arrivals.

Normalization: $\displaystyle\int_0^\infty dt\, f(t) = 1.$

Probability that the interval is larger than t: $\displaystyle P_0(t) = \int_t^\infty dt' f(t').$

Mean time interval between arrivals: $\displaystyle \tau_B = \int_0^\infty dt\, t f(t) = \int_0^\infty dt P_0(t).$

Find the probability $Q_0(t)$ that no arrivals occur in a randomly chosen time interval of length t.

First consider the probability $P_0(t'+t)$ for this to be the case if the interval starts at time t' after the last bus arrival. Then average $P_0(t'+t)$ over the range of elapsed time t'.

$$\Rightarrow\ Q_0(t) = c\int_0^\infty dt' P_0(t'+t) \text{ with normalization } Q_0(0) = 1.$$

$$\Rightarrow\ Q_0(t) = \frac{1}{\tau_B}\int_t^\infty dt' P_0(t').$$

Passengers come to the station at random times. The probability that a passenger has to wait at least a time t before the next bus is then $Q_0(t)$:

Probabilty distribution of passenger waiting times:

$$g(t) = -\frac{d}{dt}Q_0(t) = \frac{1}{\tau_B}P_0(t).$$

Mean passenger waiting time: $\displaystyle \tau_P = \int_0^\infty dt\, t g(t) = \int_0^\infty dt Q_0(t).$

The relationship between τ_B and τ_P depends on the distribution $f(t)$. In general, we have $\tau_P \leq \tau_B$. The equality sign holds for the exponential distribution.

[nex22] Pascal distribution.

Consider the quantum harmonic oscillator in thermal equilibrium at temperature T. The energy levels (relative to the ground state) are $E_n = n\hbar\omega$, $n = 0, 1, 2, \ldots$

(a) Show that the system is in level n with probability

$$P(n) = (1-\gamma)\gamma^n, \quad \gamma = \exp(-\hbar\omega/k_BT).$$

$P(n)$ is called *Pascal* distribution or *geometric* distribution.

(b) Calculate the factorial moments $\langle n^m \rangle_f$ and the factorial cumulants $\langle\langle n^m \rangle\rangle_f$ of this distribution.

(c) Show that the Pascal distribution has a larger variance $\langle\langle n^2 \rangle\rangle$ than the Poisson distribution with the same mean value $\langle n \rangle$.

Solution: