Компьютерная обработка результатов измерений

Лекция 1. Общие сведения об измерениях. Виды сигналов и методы их анализа.
Лекция 2. Статистика и вероятность. Случайные величины и распределения

Емельянов Эдуард Владимирович

Специальная астрофизическая обсерватория РАН
Лаборатория обеспечения наблюдений

18 марта 2021 года

Физические измерения

Экспериментальное определение значения измеряемой величины с применением средств измерений называется **измерением**.

Важнейшей особенностью измерений является *принципиальная невозможность получения результатов измерения, в точности равных истинному значению измеряемой величины* (особенно эта особенность проявляется в микромире, где господствует принцип неопределенности). Эта особенность приводит к необходимости оценки степени близости результата измерения к истинному значению измеряемой величины, т.е. вычислять **погрешность измерения**.

Виды измерений

Статическими называют такие измерения, при которых зависимость погрешности измерения от скорости измерения пренебрежимо мала и ее можно не учитывать. **Динамические** измерения противоположны статическим.

Результаты **прямых** измерений находят непосредственно из опыта, **косвенных** же измерений — путем расчета по известной зависимости измеряемой величины от величин, находимых прямыми измерениями (например, измерение мощности).

Совместное измерение — одновременное измерение нескольких неодноименных величин для нахождения зависимости между ними (например, ВАХ диода).

Совокупное измерение — это проведение ряда измерений нескольких величин одинаковой размерности в различных сочетаниях с нахождением искомых величин из решения системы уравнений (например, измерение R включенных треугольником резисторов).

Представление результатов

Табличное

Позволяет избежать многократной записи единиц измерения, обозначений измеряемой величины, используемых множителей. В таблицы, помимо основных измерений, могут быть включены и результаты промежуточных измерений.
Для удобства импортирования данных и одновременно наглядности чтения удобно хранить в формате TSV (tab separated values) или CSV (comma separated values). SED позволит легко преобразовать TSV/CSV в таблицу LaTeX.

Графическое

На основе графика легко можно сделать вывод о соответствии теоретических представлений данным эксперимента, определить вид функциональной зависимости измеряемой величины.

Представление результатов

Таблица 3.4: Зависимость спектрального разрешения от геометрии прибора.

θ	α	β	$\cos\alpha$	$\cos\beta$	M	$\cos\theta$	D	B	L_b
11	75.3	53.3	0.2537	0.5976	0.4245	0.9816	0.7122	0.5960	64.8
10	74.3	54.3	0.2706	0.5835	0.4637	0.9848	0.7318	0.6336	59.3
9	73.3	55.3	0.2874	0.5693	0.5048	0.9877	0.7523	0.6710	
8	72.3	56.3	0.3040	0.5548	0.5479	0.9903	0.7740	0.7079	
7	71.3	57.3	0.3206	0.5402	0.5935	0.9925	0.7966	0.7450	
6	70.3	58.3	0.3371	0.5255	0.6415	0.9945	0.8205	0.7818	
5	69.3	59.3	0.3535	0.5105	0.6925	0.9962	0.8461	0.8185	
4	68.3	60.3	0.3697	0.4955	0.7461	0.9976	0.8730	0.8546	
3	67.3	61.3	0.3859	0.4802	0.8036	0.9986	0.9017	0.8912	
2	66.3	62.3	0.4019	0.4648	0.8647	0.9994	0.9323	0.9275	
1	65.3	63.3	0.4179	0.4493	0.9301	0.9998	0.9649	0.9639	

Представление результатов

Таблица 2. Спектрографы скрещенной дисперсии в фокусе Кассегрена. Обозначения: D – диаметр телескопа; d – диаметр коллимированного пучка; θ_b – угол блеска; disp – последовательность диспергирующих элементов по ходу лучей (ech – эшелле, gr – решетка, pr – призма, filt – фильтр); R – спектральное разрешение); Obs – обсерватория. *) копии спектрографа Harvard Coll. Obs., использовавшегося на телескопе $D = 1.52$ м)

Год	**D (м)**	**d (см)**	tg θ_b	**disp**	**R**	**Obs**
1971	0.9	5.5	2	ech/gr	16000	Pine Bluff Obs. [175]
1976	0.91	5	2	pr/ech	40000	Goddard SFC [119]
1977	0.61	9	2	ech/gr	43000	Mt. John Obs. [80]
1978	0.9		2	pr/ech/pr	40000	Royal Greenwich [121]
1978	1.0	*	2	ech/gr	52000	Ritter Obs. [107]
1980	1.0	*	2	ech/gr	52000	Lowell Obs. [107]
1980	1.0		2	ech/gr	30000	Siding Spring Obs.
1981	1.0	7.7	2	ech/gr	54000	Vienna Obs. [196]
1982	0.61	*	2	ech/gr		Las Campanas [107]
1982	0.61	5	3.2	filt/ech	150000	Whipple Obs. [87]
1986	1.22		2	ech/gr	50000	Rangapur Obs.

Представление результатов

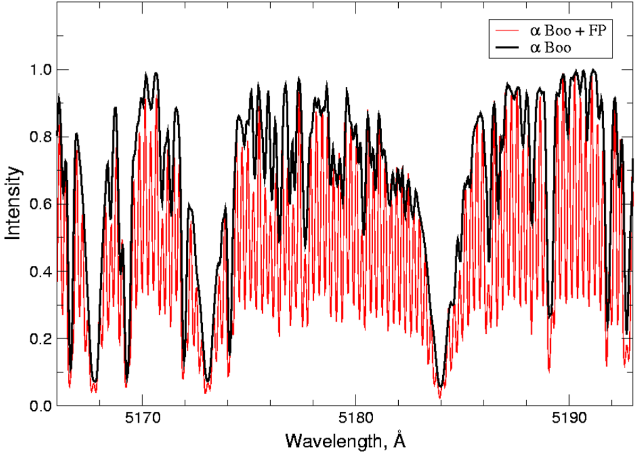


Рис. 4.7: Спектр Арктура с использованием эталона Фабри-Перо (тонкая линия) и без его применения (жирная линия).

Представление результатов

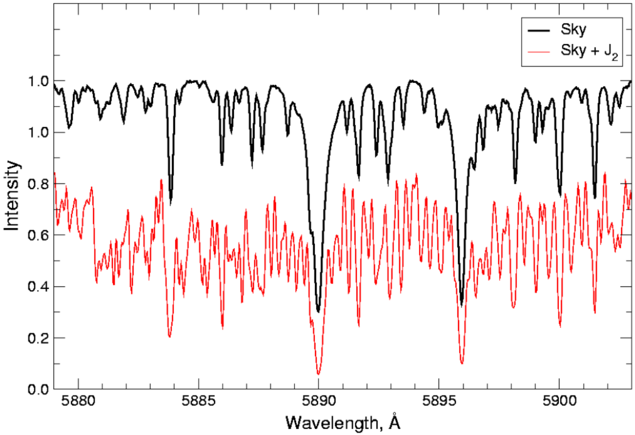


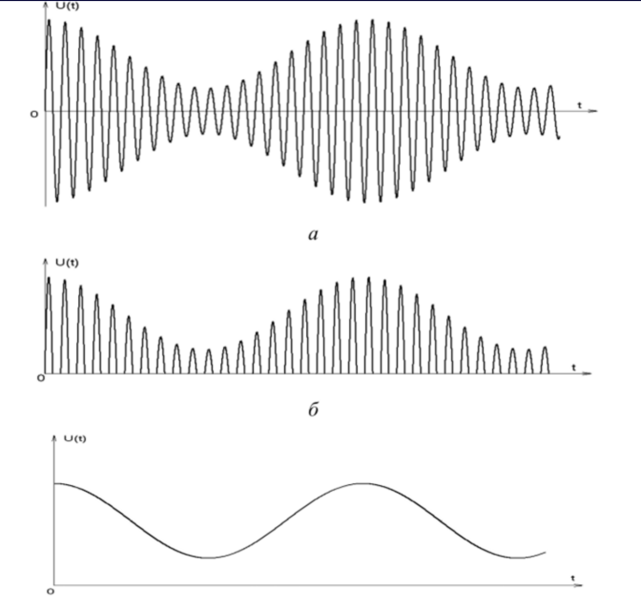
Рис. 4.8: Спектр неба с использованием абсорбционной ячейки на парах йода (тонкая линия) и без нее (жирная линия).

Сигналы и их виды

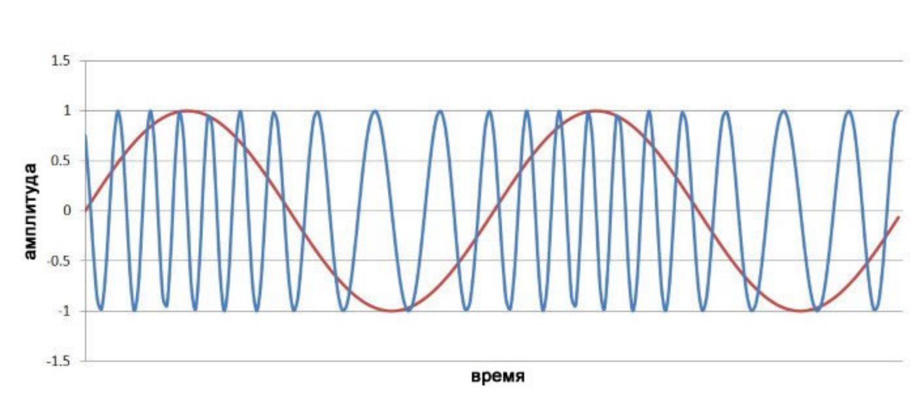
Если некоторая изменяющаяся величина измеряется непрерывно (или квазинепрерывно), мы имеем дело с потоком информации, или **сообщением**. В теории информации физический процесс, значения параметров которого отображают передаваемое сообщение, называется **сигналом**.

Модуляция–демодуляция. Зашумление. **Помехи**: аддитивные, мультипликативные, фазовые.

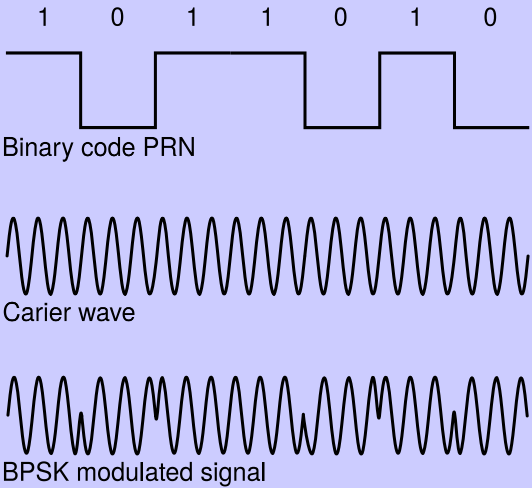
Сигналы и их виды



Сигналы и их виды

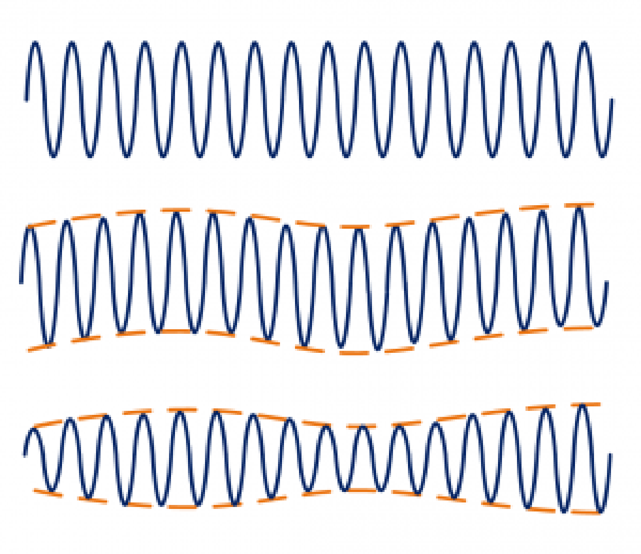


Сигналы и их виды



Сигналы и их виды

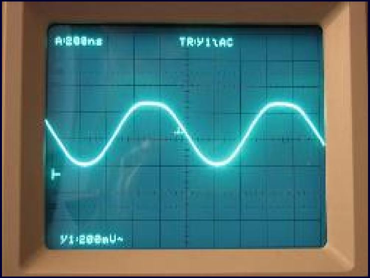
Add/mult



Виды сигналов

Аналоговый

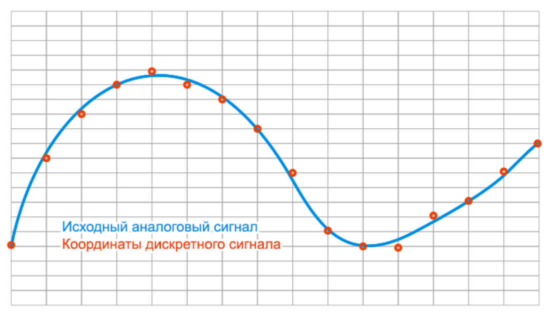
Описывается непрерывной (или кусочно–непрерывной) функцией $x(t)$: $t \in [t_0, t_1]$, $x \in [x_0, x_1]$. Аудиосигналы, телевизионные сигналы и т.п.



Виды сигналов

Дискретный

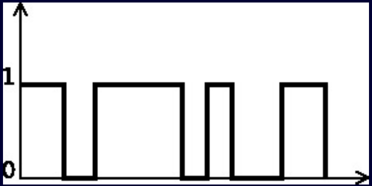
Описывается решетчатой функцией (последовательностью, временны́м рядом) $x(nT)$: $x \in [x_0, x_1]$, $n = \overline{1, N}$, T – *интервал дискретизации*. Величину $f = 1/T$ называют *частотой дискретизации*. Если интервал дискретизации является постоянной величиной, дискретный сигнал можно задать в виде ряда $\{x_1, \ldots, x_N\}$.



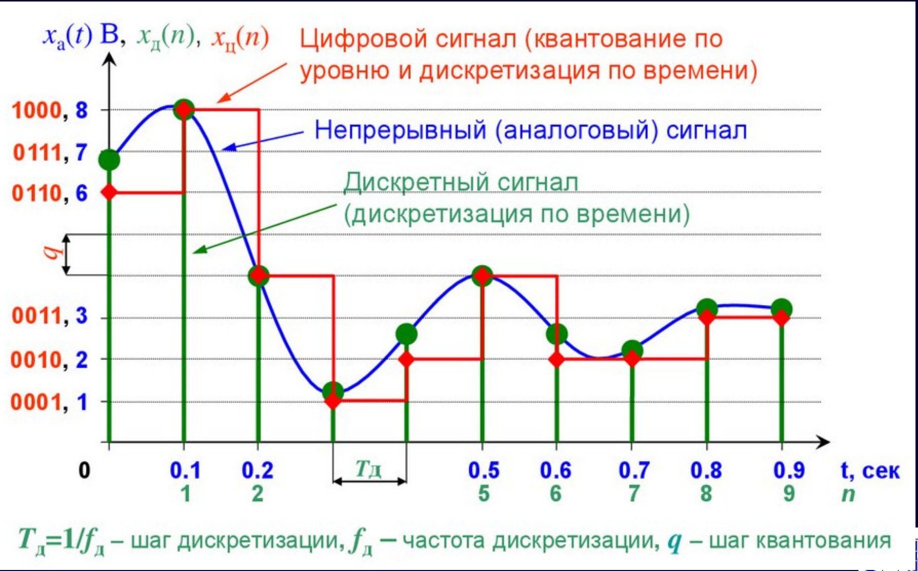
Виды сигналов

Цифровой

Описывается квантованной решетчатой функцией и отличается от обычного дискретного сигнала тем, что каждый уровень квантования кодируется двоичным кодом. Таким образом, если величина $x \in [x_0, x_1]$ квантуется N разрядным кодом, для обратного представления из кода K_x в значение x применяется преобразование: $x = x_0 + K_x \cdot (x_1 - x_0)/2^N$. К цифровым сигналам относятся сигналы, используемые в системах связи с импульсно–кодовой модуляцией.



Виды сигналов



Дискретизация

Дискретизация строит по заданному аналоговому сигналу $x(t)$ дискретный сигнал $x_n(nT)$, причем $x_n(nT) = x(nT)$. Операция **восстановления** состоит в том, что по заданному дискретному сигналу строится аналоговый сигнал.

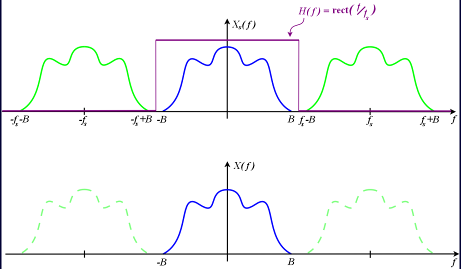
Теорема Котельникова–Найквиста

- любой аналоговый сигнал может быть восстановлен с какой угодно точностью по своим дискретным отсчётам, взятым с частотой $f > 2f_c$, где f_c – максимальная частота, которой ограничен спектр реального сигнала;
- если максимальная частота в сигнале равна или превышает половину частоты дискретизации (наложение спектра), то способа восстановить сигнал из дискретного в аналоговый без искажений не существует.

Теорема Котельникова–Найквиста

Фурье: $X_s(f) \stackrel{\text{def}}{=} \sum\limits_{n=-\infty}^{\infty} T\cdot x(nT)\, e^{-i2\pi nTf}$

В окне: $X(f) = \sum\limits_{n=-\infty}^{\infty} x(nT)\cdot \underbrace{T\cdot \mathrm{rect}(Tf)\cdot e^{-i2\pi nTf}}_{\mathcal{F}\left\{\mathrm{sinc}\left[\frac{\pi}{T}(t-nT)\right]\right\}}$



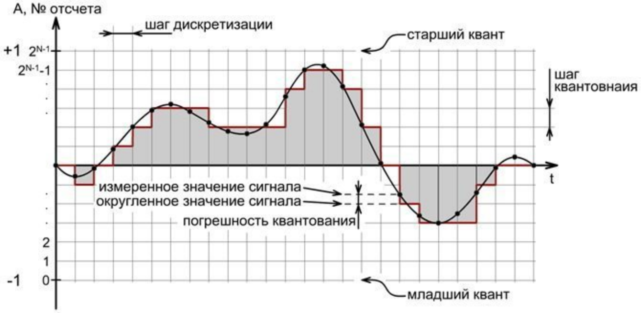
Формула Уиттекера–Шеннона

Восстановить непрерывную функцию из дискретной:

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT)\cdot \mathrm{sinc}\left[\frac{\pi}{T}(t-nT)\right]$$

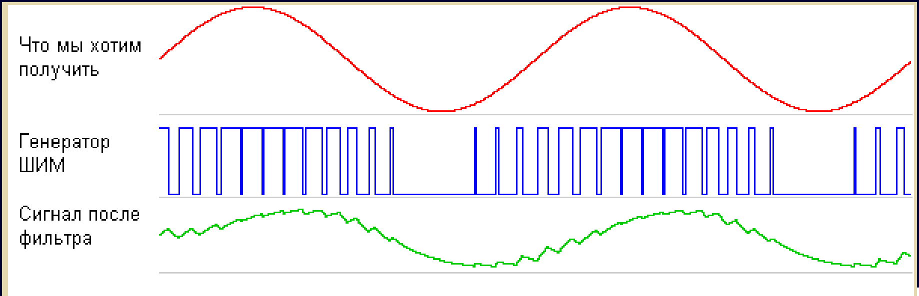
Квантование

Для преобразования дискретного сигнала в цифровой вид применяется операция **квантования** или **аналогово–цифрового преобразования** (АЦП), которая по заданному дискретному сигналу $x_n(nT)$ строит цифровой кодированный сигнал $x_d(nT)$, причем $x_n(nT) \approx x_d(nT)$. Обратная квантованию операция называется операцией **цифро–аналогового преобразования** (ЦАП).



Квантование

Для преобразования дискретного сигнала в цифровой вид применяется операция **квантования** или **аналогово–цифрового преобразования** (АЦП), которая по заданному дискретному сигналу $x_n(nT)$ строит цифровой кодированный сигнал $x_d(nT)$, причем $x_n(nT) \approx x_d(nT)$. Обратная квантованию операция называется операцией **цифро–аналогового преобразования** (ЦАП).



Основная литература

- Интернет–энциклопедия: http://wikipedia.org (Википедия).
- Гонсалес Р., Вудс Р. Цифровая обработка изображений. — М.: Техносфера, 2012. — 1104 с.
- Витязев В.В. Вейвлет-анализ временных рядов: Учеб. пособие. — СПб.: Изд-во С.-Петерб. ун-та., 2001. — 58 с.
- Гонсалес Р., Вудс Р., Эддинс С. Цифровая обработка изображений в среде MATLAB. — М.: Техносфера, 2006 — 616 с.
- Гмурман В. Е. Теория вероятностей и математическая статистика. Учеб. пособие для вузов. — Изд. 7-е, стер. — М.: Высш. шк., 2001. — 479 с.
- Говорухин В., Цибулин В. Компьютер в математическом исследовании. Учебный курс. — СПб.: Питер, 2001. — 624 с.
- Сергиенко А. Б. Цифровая обработка сигналов. — СПб.: Питер, 2005. — 604 с.
- Чен К., Джиблин П., Ирвинг А. MATLAB в математических исследованиях: Пер. с англ. — М.: Мир, 2001. — 346 с.

Дополнительная литература

- Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. — М.: Высш. шк., 1987. — 630 с.
- Кнут Д. Э. Все про TEX./ Пер. с англ. М. В. Лисиной. — Протвино: АО RDTEX, 1993. — 592 с.: ил.
- Львовский С. М. Набор и верстка в системе LATEX. — 3-е изд., испр. и доп. — М.: МЦНМО, 2003. — 448 с.
- Физическая энциклопедия/ Гл. ред. А.М. Прохоров. — М.: Сов. энциклопедия. Тт. I – V. 1988.
- Цифровая обработка сигналов: Справочник/ Л.М. Гольденберг, Б.Д. Матюшкин, М.Н. Поляк. — М.: Радио и связь, 1985. — 312 с., ил.
- http://www.imageprocessingplace.com/
- Pan G. W. Wavelets in electromagnetic and device modeling. — John Wiley & Sons, Inc., Hobocen, New Jersey, 2003. — 531 p.

Лекция 2.

Случайные величины, вероятность

Случайной величиной называется величина X, если все ее возможные значения образуют конечную или бесконечную последовательность чисел $x_1, \ldots, x_N$, и если принятие ею каждого из этих значений есть случайное событие.

Вероятностью наступления события называют предел относительной частоты наступления данного события n_k/N:

$$P(x_k) = \lim_{N\to\infty} \frac{n_k}{N}.$$

Совместные и несовместные события, полная группа, свойства вероятности. Для непрерывных случайных величин вводят понятие **плотности вероятности**:

$$\rho(x) = \lim_{\Delta x\to 0} \frac{P(x < X < x + \Delta x)}{\Delta x} = \frac{dP}{dx}.$$

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} \rho(x)\, dx.$$

Свойства вероятности

$P(\emptyset) = 0$	
$\forall A \subset B \quad P(A) \leqslant P(B)$	B включает в себя A
$0 \leqslant P(A) \leqslant 1$	
$\forall A \subset B \quad P(B \setminus A) = P(B) - P(A)$	B наступит без A
$P(\overline{A}) = 1 - P(A)$	
$P(A+B) = P(A) + P(B) - P(AB)$	вероятность одного из событий
$P(A\|B) = {P(AB)}/{P(B)}$	условная вероятность (A при B) $\Longrightarrow$
$P(AB) = P(B) \cdot P(A\|B)$	или $P(AB) = P(A) \cdot P(B\|A) \Longrightarrow$
$P(A\|B) = \dfrac{P(A) \cdot P(B\|A)}{P(B)}$	(теорема Байеса)
$P(AB) = P(A) \cdot P(B)$	для независимых событий

Характеристики случайных величин

Среднее арифметическое и математическое ожидание

$$<X> = {}^1\!/_N \sum_{n=1}^{N} x_n,$$

$$M(X) \equiv \overline{X} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n \quad \text{и} \quad M(X) = \int\limits_{-\infty}^{\infty} x\varphi(x)\,dx.$$

Свойства математического ожидания

- $\overline{\mathfrak{const}} = \mathfrak{const}$;
- $\overline{\sum \mathfrak{C}_n \cdot X_n} = \sum \mathfrak{C}_n \cdot \overline{X_n}$, где $\mathfrak{C}_n$ – постоянная величина;
- $\overline{\prod X_n} = \prod \overline{X_n}$ (для независимых случайных величин);
- $\overline{f(x)} = \int\limits_{-\infty}^{\infty} f(x)\varphi(x)\,dx$ (для непрерывных случайных величин).

Моменты

Если $f(x) = (x - x_0)^n$, то $\overline{f(x)}$ — момент порядка n. Если $x_0 = 0$ — начальный момент, если $x_0 = \overline{X}$— центральный момент.
Центральный момент второго порядка называют **дисперсией**:
$D(X) = \overline{(x - \overline{x})^2} \equiv \overline{x^2} - \overline{x}^2$, $\sigma = \sqrt{D}$.

Свойства дисперсии:

- $D(\mathfrak{C}) = 0$;
- $D(\mathfrak{C}X) = \mathfrak{C}^2 D(X)$, где $\mathfrak{C}$ – постоянная величина;
- $D(\sum X_n) = \sum D(X_n)$ (для независимых величин).

$\overline{X} \Leftrightarrow <X>$? Закон больших чисел

Неравенство Чебышёва: $P(|X - \overline{X}| \geqslant \varepsilon) \leqslant {D(X)}/{\varepsilon^2} \quad \Rightarrow$
$P(|X - \overline{X}| < \varepsilon) = 1 - P(|X - \overline{X}| \geqslant \varepsilon) \geqslant 1 - {D(X)}/{\varepsilon^2}$.

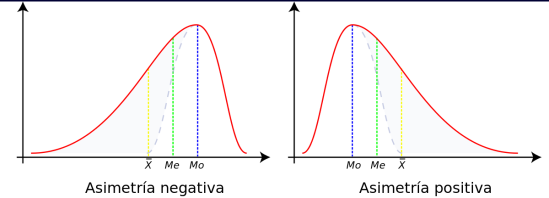
$$\lim_{n\to\infty} P\left(\left|\frac{\sum X_n}{n} - \frac{\sum \overline{X_n}}{n}\right| < \varepsilon\right) = 1, \quad \text{т.к.} \quad D(\textstyle\sum X_n/n) = D(X)/n$$

Теорема Бернулли: $\lim\limits_{n\to\infty} P(m/n - p| < \varepsilon) = 1$ (m событий в n испытаний).

Характеристические значения распределений

Медиана и мода

Мода — наиболее часто встречающееся значение (но вполне могут быть мультимодальные распределения). **Медиана** делит площадь распределения пополам.



Поиск медианы

Самый медленный — сортировкой ряда данных, $O(n \ln n)$. Quick Select, $O(n)$. Гистограмма (в т.ч. дерево гистограмм), $O(n)$. Для фиксированных n — opt_med („Numerical Recipes in C“), $O(n)$.

Законы распределения

Закон распределения *дискретной* случайной величины — соответствие между возможными значениями и их вероятностями.

Функция распределения

$$F(x) \equiv P(X \leqslant x) = \int\limits_{-\infty}^{x} \varphi(x)\,dx, \qquad \int\limits_{-\infty}^{\infty} \varphi(x)\,dx = 1.$$

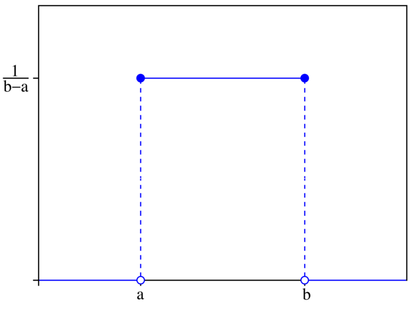
$$P(a \leqslant X \leqslant b) = F(b) - F(a).$$

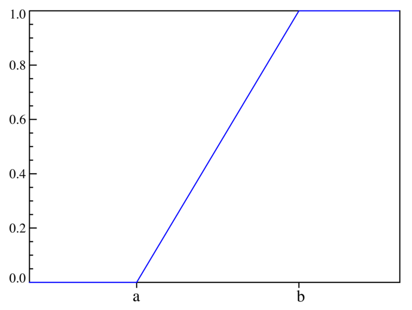
Равномерное распределение

$$\varphi(x)=\begin{cases}\frac{1}{b-a}, & x\in[a,b]\\ 0, & x\notin[a,b]\end{cases}.$$

$$F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\leqslant x<b\\ 1, & x\geqslant b\end{cases}.$$

$$\overline{X}=\mathrm{med}(X)=(a+b)/2,$$
$$\mathrm{Mo}(X)=\forall x\in[a,b],$$
$$\sigma_X^2=\frac{(b-a)^2}{12}.$$



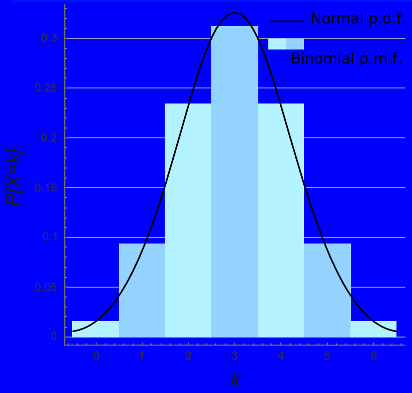


Биномиальное распределение

Формула Бернулли: $P_n(k) = C_n^k p^k q^{n-k}, \quad C_n^k = \dfrac{n!}{k!(n-k)!}, \quad q = 1 - p.$

$$(p+q)^n = C_n^n p^n + \cdots + C_n^k p^k q^{n-k} + \cdots + C_n^0 q^k.$$

Описывает вероятность наступления события k раз в n независимых испытаниях



$$F(k;n,p) = P(X \le k) = \sum_{i=0}^{\lfloor k \rfloor} C_n^i p^i (1-p)^{n-i}.$$

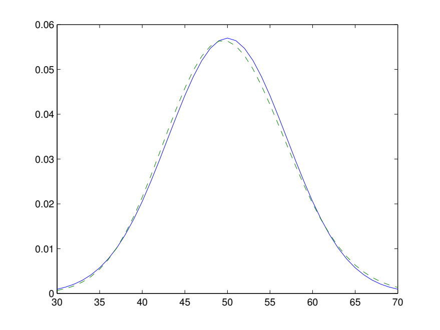
$\overline{X} = np$, $\mathrm{Mo}(X) = \lfloor (n+1)p \rfloor$,
$\lfloor np \rfloor \leqslant \mathrm{med}(X) \leqslant \lceil np \rceil$, $\sigma_X^2 = npq$.

Распределение Пуассона

При $n \to \infty$ распределение Бернулли преобразуется в распределение Пуассона ($\lambda = np$):

$$P_n(k) = \frac{\lambda^k}{k!}\exp(-\lambda).$$

$F(k,\lambda) = \dfrac{\Gamma(k+1,\lambda)}{k!}$, $\overline{X} = \lambda$,
$\mathrm{Mo}(X) = \lfloor\lambda\rfloor$,
$\mathrm{med}\,X \approx \lfloor\lambda + 1/3 - 0.02/\lambda\rfloor$,
$\sigma_X^2 = \lambda$.
С ростом λ распределение Пуассона стремится к распределению Гаусса.

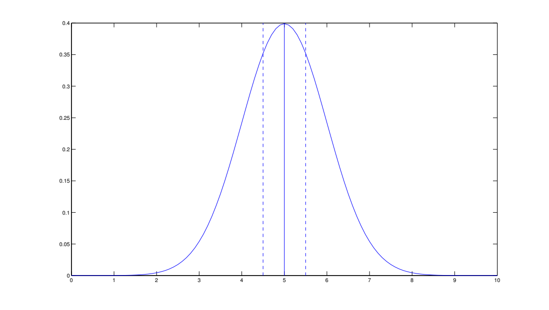


Распределение Гаусса

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\tfrac{(x-\overline{x})^2}{2\sigma^2}\right),\ F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left(-\frac{(t-\overline{x})^2}{2\sigma^2}\right)\,dt,$$

$\mathrm{Mo}(X) = \mathrm{med}\,X = \overline{X}$. $P(\alpha < X < \beta) = \Phi\left(\frac{\beta-\overline{x}}{\sigma}\right) - \Phi\left(\frac{\alpha-\overline{x}}{\sigma}\right)$,

функция Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left(-t^2/2\right)$.

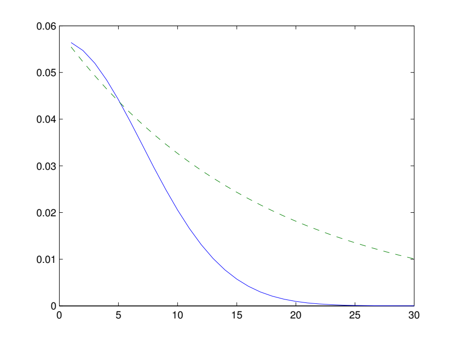


Показательное (экспоненциальное) распределение

Время между двумя последовательными свершениями события

$$f(x)=\begin{cases}0, & x<0,\\ \lambda\exp(-\lambda x), & x\geqslant 0;\end{cases}\qquad F(x)=\begin{cases}0, & x<0,\\ 1-\exp(-\lambda x), & x\geqslant 0,\end{cases}$$

$\overline{X}=\lambda^{-1}$, $\mathrm{Mo}(X)=0$, $\mathrm{med}\,X=\ln(2)/\lambda$, $\sigma_X^2=\lambda^{-2}$.



Корреляция и ковариация

Ковариация является мерой линейной зависимости случайных величин и определяется формулой: $\mathrm{cov}(X,Y) = \overline{(X-\overline{X})(Y-\overline{Y})} \Longrightarrow \mathrm{cov}(X,X) = \sigma_X^2$. *Ковариация независимых случайных величин равна нулю*, обратное неверно.

Если ковариация положительна, то с ростом значений одной случайной величины, значения второй имеют тенденцию возрастать, а если знак отрицательный — убывать.
Масштаб зависимости величин пропорционален их дисперсиям $\Longrightarrow$ масштаб можно отнормировать (**коэффициент корреляции** Пирсона):

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma X \sigma Y}, \qquad \mathbf{r} \in [-1,1].$$

Коэффициент корреляции равен ± 1 тогда и только тогда, когда X и Y линейно зависимы. Если они независимы, $\rho_{X,Y} = 0$ (**обратное неверно!**). Промежуточные значения коэффициента корреляции не позволяют однозначно судить о зависимости случайных величин, но позволяет предполагать степень их зависимости.

Корреляционная функция

Одна из разновидностей — **автокорреляционная функция**:

$$\Psi(\tau) = \int f(t) f(t-\tau)\, dt \equiv \int f(t+\tau) f(t)\, dt.$$

Для дискретных случайных величин автокорреляционная функция имеет вид

$$\Psi(\tau) = <X(t)X(t-\tau)> \equiv <X(t+\tau)X(t)>.$$

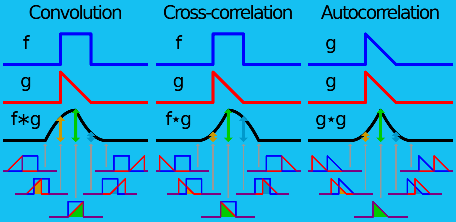
Взаимно корреляционная функция

Другая разновидность — **кросс–корреляционная функция**:

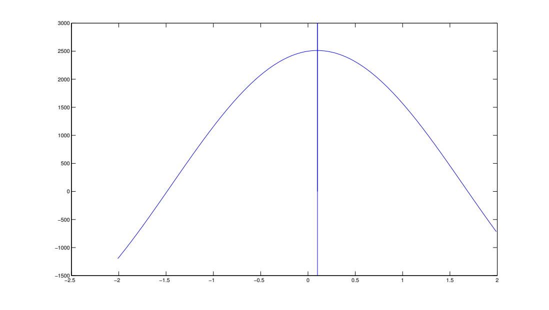
$$(f \star g)(t) \stackrel{\text{def}}{=} \int\limits_{-\infty}^{\infty} f^*(\tau) g(t+\tau)\, d\tau$$

свертка:

$$(f * g)(x) \stackrel{\text{def}}{=} \int\limits_{-\infty}^{\infty} f(y)\, g(x-y)\, dy = \int\limits_{-\infty}^{\infty} f(x-y)\, g(y)\, dy.$$



Если X и Y — две независимых случайных величины с функциями распределения вероятностей f и g, то $f \star g$ соответствует распределению вероятностей выражения $-X+Y$, а $f * g$ — распределению вероятностей суммы $X+Y$.
ВКФ часто используется для поиска в длинной последовательности более короткой заранее известной, определения сдвига (см. рис).
Связь со сверткой: $f(t) \star g(t) = f^*(-t) * g(t)$, если f и g четны, то $f(t) \star g(t) = f(t) * g(t)$. Через преобразование Фурье: $\mathcal{F}(f \star g) = \mathcal{F}(f)^* \cdot \mathcal{F}(g)$.



Шум

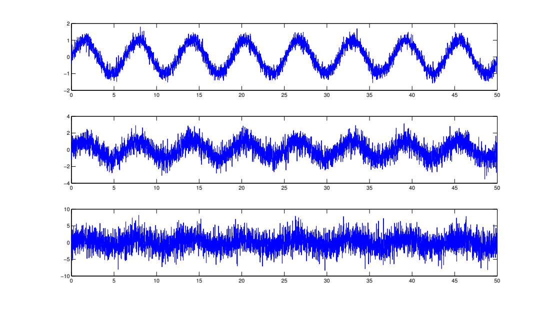
Шум — беспорядочные колебания различной физической природы, отличающиеся сложной временной и спектральной структурой.

Белый шум, $\xi(t)$, имеет время корреляции много меньше всех характерных времен физической системы; $\overline{\xi(t)} = 0$, $\Psi(t,\tau) = <\xi(t+\tau)\xi(t)> = \sigma^2(t)\delta(\tau)$. Разновидность — AWGN.
Дробовой шум имеет пуассонову статистику $\Longrightarrow \sigma_X \propto \sqrt{x}$ и $\mathrm{SNR}(N) \propto \sqrt{N}$. Суточные и вековые корреляции.
Шум вида **«соль–перец»** обычно характерен для изображений, считываемых с ПЗС.

SNR

SNR — безразмерная величина, равная отношению мощности полезного сигнала к мощности шума.

$$\mathrm{SNR} = \frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}} = \left(\frac{A_{\mathrm{signal}}}{A_{\mathrm{noise}}}\right)^2, \quad \mathrm{SNR}(dB) = 10\lg\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right) = 20\lg\left(\frac{A_{\mathrm{signal}}}{A_{\mathrm{noise}}}\right).$$



(10, 0, -10 дБ.)

Спасибо за внимание!

mailto

eddy@sao.ru
edward.emelianoff@gmail.com