

Определение вероятностей сложных событий

От простых событий к сложному

Вычисление вероятностей сложных событий сводится к применению формул сложения и умножения вероятностей, если вероятности всех составляющих простых событий известны. В объектно-ориентированной среде вероятность события, выраженного через суммы и произведения простых событий, вычисляется автоматически методами класса `Randev`.

Вероятностная модель испытаний Бернулли

В численном эксперименте частота моделируемого события практически перестает изменяться от серии к серии при пятизначном числе повторений в серии (см. рис. 1.2, а). Объем физических испытаний не может быть таким большим, но должен быть достаточным для получения требуемой точности оценки вероятности. Каким должно быть число испытаний n , чтобы частота m/n наступления события не отличалась от вероятности p больше, чем на ε ?

Теорема Бернулли

Вопрос правомерен, но некорректно сформулирован. Ни при каком n нельзя гарантировать выполнение неравенства $|m/n - p| \leq \varepsilon$, потому что это случайное событие (не исключено ни одно значение m от 0 до n). Можно лишь потребовать, чтобы оно выполнялось с большой вероятностью, превышающей некоторый (доверительный) уровень β (0,9, 0,95 или больше):

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \geq \beta = 1 - \alpha. \quad (2.1)$$

При неограниченном числе повторений *независимых опытов* (исход любого из них не зависит от результатов других опытов) неравенство (2.1) выполняется для сколь угодно малых ε и α (теорема Я. Бернулли).

Формула Бернулли

Испытаниями Бернулли называют последовательность n независимых опытов с двумя возможными исходами в каждом (успех A , неудача \bar{A}) и одинаковой вероятностью успеха p в каждом повторении. Каждая последовательность из m успехов и $n - m$ неудач представляет собой произведение m событий A и $n - m$ событий \bar{A} . В силу независимости вероятность их произведения равна произведению вероятностей $p^m(1 - p)^{n - m}$. Событие $A_{m,n}$ наступления ровно m успехов все равно, в какой очередности успехов и неудач, представляет собой сумму C_n^m несовместных событий с одинаковыми вероятностями $p^m(1 - p)^{n - m}$. Таким образом, вероятность ровно m успехов в n испытаниях Бернулли определяется *биномиальной формулой*:

$$P(A_{m,n}) \equiv P_{m,n} = C_n^m p^m (1 - p)^{n - m}. \quad (2.2)$$

Обобщенная формула Бернулли

Требование бинарности исходов (успех или неудача) необязательно. Если в каждом из n независимых испытаний может наступить одно из событий A_1, \dots, A_k с вероятностями $P(A_i) = p_i$, ($\sum p_i = 1$), вероятность события A_{m_1, \dots, m_k} (m_1 раз наступит A_1 , m_2 раз — A_2 и т.д., $\sum m_i = n$), определяется по формуле:

$$P(A_{m_1, \dots, m_k}) = \frac{n!}{m_1! \dots m_k!} p_1^{m_1} \dots p_k^{m_k}. \quad (2.3)$$

**Практическое
нахождение
достаточного
числа повторений**

Для практических вычислений по формулам (2.2), (2.3) можно составить функцию, которую удобно использовать для разных схем независимых испытаний. Аргументы функции $p_Bern(P, n, M)$ (Листинг 2.1) позволяют организовать вычисления как по биномиальной, так и по обобщенной формуле. Первый аргумент задает вероятности p_1, \dots, p_k в обозначениях формулы (2.3), но достаточно задать только первые $k-1$ элементов этого вектора, последний будет вычислен как дополнение до 1. Так же задаются числа успехов m_1, \dots, m_k (или m_1, \dots, m_{k-1}) третьим аргументом. При необходимости последний элемент m_k вычисляется как дополнение до числа испытаний, задаваемого вторым аргументом. Благодаря контролю аргументов вызов функции $p_Bern(p, n, m)$ со скалярными аргументами возвращает $P_{m,n}$ в обозначениях формулы (2.2). Такой же результат даст и функция $p_Binom(p, n, m)$ (Листинг 2.2), специализирующаяся на вычислениях по формуле (2.2). Эта функция удобна тем, что она векторизована по третьему аргументу, которым можно определить вектор интересующих чисел успехов (например, не менее трех 3:n, не более трех 0:3 или всех возможных чисел успехов 0:n). Аргумент 0:n предполагается по умолчанию (его можно опустить). Функция возвращает вероятности каждого из чисел успехов, поэтому вероятность, например, не менее трех успехов можно получить, применяя операцию суммирования к результату p_Binom : $\text{sum}(p_Binom(p, n, 3:n))$.

С помощью электронной формулы p_Binom можно найти необходимое число опытов по условию (2.1). Выбрав заведомо недостаточное значение n , установим диапазон допустимых значений числа успехов $[m_1, m_2]$ при данном ϵ так, что $m_1 = n(p - \epsilon)$, $m_2 = n(p + \epsilon)$ с округлением m_1 до большего, а m_2 до меньшего целого. Вычислим $P_{m,n}$ для всех m из этого диапазона и, если сумма этих вероятностей меньше доверительной вероятности 0,9, увеличим n и повторим процедуру:

```
>> p=0.75;eps=0.1; p1=p-eps;p2=p+eps; Pk=0.9;
>> for n=9:90 m=ceil(n*p1):fix(n*p2); P=sum(Ber (p,n,m));if P>Pk break,end,end,n,P
n = 40 P = 0.9023
```

Вероятность попадания частоты в допустимый интервал $[0,65, 0,85]$ превышает доверительную $0,9023 > 0,9$ при $n = 40$ повторениях опыта.

Существует аппроксимация биномиальной формулы, позволяющая не только избежать многократных вычислений факториалов, но и в явном виде получить зависимость между точностью, доверительной вероятностью и необходимым объемом испытаний, которую можно использовать для качественной организации статистических испытаний.

**Локальная теорема
Лапласа**

Если вероятность успеха в каждом опыте постоянна и отличается от 0 и 1 ($0 < p < 1$), то вероятность $p_{m,n}$ того, что успех наступит ровно m раз, приближенно (тем точнее, чем больше n) равна

$$p_{m,n} \approx \frac{1}{\sqrt{npq}} \varphi(x), \quad (2.4)$$

где $q = 1 - p$,

$$x = \frac{m - np}{\sqrt{npq}}, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (2.5)$$

Аппроксимация биномиального распределения непрерывной функцией позволяет получить интегральную формулу для определения суммарной вероятности возможных значений СВ в заданном интервале.

**Интегральная теорема
Лапласа**

В предположениях локальной теоремы вероятность того, что в n испытаниях успех наступит не менее m_1 и не более m_2 раз приближенно равна:

$$P_n(m_1 \leq m \leq m_2) \approx \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x^2}{2}} dx \quad (2.6)$$

где,

$$x_1 = \frac{m_1 - np}{\sqrt{npq}}, x_2 = \frac{m_2 - np}{\sqrt{npq}}. \quad (2.7)$$

В справочниках имеются таблицы значений функции Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \quad (2.8)$$

для неотрицательных аргументов. Функция Лапласа нечетна, поэтому определенный интеграл в формуле (2.6) можно выразить следующим образом

$$\frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{x_1}^0 e^{-\frac{x^2}{2}} dx + \frac{1}{\sqrt{2\pi}} \int_0^{x_2} e^{-\frac{x^2}{2}} dx = \Phi(x_2) - \Phi(x_1).$$

Условие для необходимого числа испытаний

Вероятность того, что число успехов в испытаниях Бернулли с параметрами p, n находится в интервале $[m_1, m_2]$, можно вычислить как разность значений функции Лапласа с аргументами, определяемыми по формулам (2.7):

$$x_{2,1} = \frac{n(p \pm \varepsilon) - np}{\sqrt{npq}} = \pm \varepsilon \sqrt{\frac{n}{pq}},$$

$$P(m_1 \leq m \leq m_2) = \Phi(x_2) - \Phi(x_1) = 2\Phi(x_2) = 2\Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right).$$

С помощью обратной функции Φ^{-1} свяжем условие $P(m_1 \leq m \leq m_2) > \beta$ со значением аргумента монотонно возрастающей функции Лапласа

$$\varepsilon \sqrt{\frac{n}{pq}} \geq \Phi^{-1}\left(\frac{\beta}{2}\right), \quad (2.9)$$

откуда получим искомое условие для объема испытаний:

$$n \geq \frac{pq}{\varepsilon^2} \left(\Phi^{-1}\left(\frac{\beta}{2}\right) \right)^2. \quad (2.10)$$

Вычисления функции Лапласа в среде MATLAB

Файл-функция `f_Gauss` (Листинг 2.2) вычисляет функцию (2.5) для всех элементов векторного аргумента, поэтому ее удобно использовать в процедуре численного интегрирования `Trap` (Листинг 2.3), которая применяет метод трапеций к уже вычисленному вектору значений подынтегральной функции на расчетной сетке. Она использована в электронной формуле для функции Лапласа `f_Laplas` (Листинг 2.4). С помощью векторизованной версии `f_LaplasV` (Листинг 2.5) составим в качестве примера таблицу функций $f(x)$ и $\Phi(x)$:

```
>> x=0:0.1:0.9;X=[x' 1+x' 2+x'];f=f_Gauss(X);Fi=F_LaplasV(X);
>> disp([X(:,1) f(:,1) Fi(:,1) X(:,2) f(:,2) Fi(:,2) X(:,3) f(:,3) Fi(:,3)])]
```

Таблица 2.1. Значения функций $f(x)$ и $\Phi(x)$.

x	$f(x)$	$\Phi(x)$	x	$f(x)$	$\Phi(x)$	x	$f(x)$	$\Phi(x)$
0.0000	0.3989	0.0000	1.0000	0.2420	0.3413	2.0000	0.0540	0.4772
0.1000	0.3970	0.0398	1.1000	0.2179	0.3643	2.1000	0.0440	0.4821
0.2000	0.3910	0.0793	1.2000	0.1942	0.3849	2.2000	0.0355	0.4861
0.3000	0.3814	0.1179	1.3000	0.1714	0.4032	2.3000	0.0283	0.4893
0.4000	0.3683	0.1554	1.4000	0.1497	0.4192	2.4000	0.0224	0.4918
0.5000	0.3521	0.1915	1.5000	0.1295	0.4332	2.5000	0.0175	0.4938
0.6000	0.3332	0.2257	1.6000	0.1109	0.4452	2.6000	0.0136	0.4953
0.7000	0.3123	0.2580	1.7000	0.0940	0.4554	2.7000	0.0104	0.4965
0.8000	0.2897	0.2881	1.8000	0.0790	0.4641	2.8000	0.0079	0.4974
0.9000	0.2661	0.3159	1.9000	0.0656	0.4713	2.9000	0.0060	0.4981

Решение обратной задачи

Вычисление обратной функции $\Phi^{-1}(x)$ с помощью электронной формулы ArgLaplas (Листинг 2.6) позволит решить неравенство (2.10) и подобные обратные задачи без обращения к таблице. Например, значение $\Phi^{-1}(0.9/2)$ согласно таблице 2.1 находится между 1,6 и 1,7. С помощью функции ArgLaplas его можно вычислить точнее:

```
>> x=ArgLaplas(0.9/2)
x = 1.6450
```

При тех же исходных данных $p = 0,75$, $P = 0,9$, $\varepsilon = 0,1$, которые использовались для прямых вычислений по электронной формуле p_Binom , получим:

```
>> n=3/4*1/4*(ArgLaplas(0.9/2)/0.1)^2;n=round(n),m1=round(n*[0.65 0.85])
n = 51    m1=33    m2 = 43
```

По сравнению с $n = 40$, полученным ранее, необходимый объем испытаний по нормальному приближению существенно больше. Но и вероятность события $m \in [33, 43]$ существенно больше требуемой:

```
>> P=sum(p_Binom(p,51,33:43))
P = 0.9259
```

Использование нормального приближения для аппроксимации биномиальной формулы привело к завышению левой части неравенства (2.1) и, соответственно, объема испытаний из-за того, что аппроксимация не учитывает дискретность аргументов m_1, m_2 . Если увеличить доверительную вероятность на 0,92 в точном вычислении, получится практически тот же результат, что и с использованием нормального приближения:

```
>> K=0.92;for n=10:100 a=ceil(n*p1):fix(n*p2);P=sum(Ber(p,n,a));if P>K break,end,end,n,P
n = 52    P = 0.9239
```

То, что в неравенстве (2.10) явно связаны все параметры задачи, имеет свои преимущества. Например, когда вероятность успеха неизвестна и исследования ведутся с целью выяснения ее зависимости от условий опыта, подстановка в (2.10) вместо pq его максимально возможного значения $1/4$ увеличивает необходимое число испытаний, как «плату» за отсутствие априорной информации:

```
>> n=1/2*1/2*(ArgLaplas(0.9/2)/0.1)^2
n = 67.6498
```

Это не очень ценный результат. На практике нужно по результатам проведенных опытов определять доверительный интервал для вероятностей наступления события во всем диапазоне условий (кривая чувствительности, кривая вероятности пробития и т.п.) Однако по условию локальной теоремы Лапласа нормальное приближение неприменимо, если вероятность успеха очень мала или близка к единице, а значит его нельзя использовать при определении нижней и верхней границ кривой чувствительности.

Формула Пуассона

В неограниченном числе независимых испытаний, если $n \rightarrow \infty$ и $p \rightarrow 0$ так, что $np \rightarrow \lambda = \text{const}$, вероятность того, что успех наступит ровно m раз, определяется *формулой Пуассона*:

$$p_m = \frac{\lambda^m}{m!} e^{-\lambda}, m = 0, 1, \dots \quad (2.10)$$

С помощью электронной формулы $p_Poisson$ (Листинг 2.6), векторизованной по второму аргументу (возможным значениям), вычислим вероятность того, что в $n = 100$ испытаниях устройства, вероятность безотказной работы которого $p = 0,99$, произойдет не более трех отказов, вычислим как сумму p_m , $m = 0, 1, 2, 3$:

```
>> R4=sum(p_Poisson(1,0:3))
R4 = 0.9810
```

Сравним этот результат с точным, вычислением по биномиальной формуле:

```
>> R_4=sum(p_Binom(0.01,100,0.3))
R_4 = 0.9816
```

При том же значении $np = 1$, но при малом числе повторений разница увеличится:

```
>> n=10;p=0.1;R4=sum(p_Poisson(p*n,0.3)),R_4=sum(p_Binom(p,n,0.3))
R4 = 0.9810 R_4 = 0.9872
```

Оценка вероятности по частоте успехов в испытаниях Бернулли

В испытаниях Бернулли с неизвестной вероятностью успеха p по частоте $p^* = m/n$ интересующего события можно лишь утверждать, что разность $|p^* - p|$ не превосходит ε с доверительной вероятностью β : $P(|p^* - p| < \varepsilon) > \beta$. Так, если в 10 испытаниях успех наблюдался 4 раза, это не исключает того, что $p = 0,4$. Можно исключить лишь $p > p_2$ и $p < p_1$ при p_1, p_2 таких, что 4 успеха в n испытаниях становится маловероятным событием.

Определение вероятности события по его частоте

Если частота не слишком мала и не близка к единице, для нахождения p_1, p_2 можно воспользоваться нормальным приближением вероятности попадания частоты в симметричный интервал относительно параметра p . Решим неравенство (2.9) относительно $\varepsilon \equiv |p^* - p|$:

$$|p^* - p| < \sqrt{\frac{pq}{n}} \Phi^{-1}\left(\frac{\beta}{2}\right).$$

Обозначив $t_\beta = \Phi^{-1}(\beta/2)$, представим неравенство

$$(p - p^*)^2 < p(1-p) \frac{t_\beta^2}{n}$$

в виде $Ap^2 - 2Bp + C < 0$, и получим границы доверительного интервала:

$$p_{1,2} = \frac{B \pm \sqrt{B^2 - AC}}{A},$$

где $A = 1 + \frac{t_\beta^2}{n}$, $B = p^* + \frac{t_\beta^2}{2n}$, $C = p^{*2}$.

Обычно для удобства практических применений строят кривые (эллипсы) зависимостей $p_1(p^*)$ и $p_2(p^*)$ для всего диапазона $0 < p^* < 1$ при различных n с фиксированным $\beta = 0,9$ или $\beta = 0,95$. Семейство кривых на рис. 2.3 построено при $\beta = 0,9$ и нескольких n , для чего вторая командная строка выполнена 6 раз с заменой $n = 5$ на 10, 20, 50, 100 и 500:

```
>> P=0.9;T=ArgLaplas(P/2)^2;x=0.01:0.01:0.99;C=x.^2; hold on
>> n=5; A=1+T/n;B=x+T/n/2;D=sqrt(B.^2-A*C);p2=(B+D)/A;p1=(B-D)/A;plot(x,[p2;p1])
```

Если в 50 опытах успех наблюдался 20 раз, частоте 0,4 соответствует доверительный интервал $[0,29, 0,52]$, показанный на рисунке. Выполним последнюю команду при $x=0,4$, чтобы вычислить этот же результат точнее:

```
>>n=50;x=0.4;C=x^2;t=T/n;A=1+t;B=x+t/2;D=sqrt(B.^2-A*C);p2=(B+D)/A;p1=(B-D)/A
p2 = 0.5163 p1 = 0.2940
```

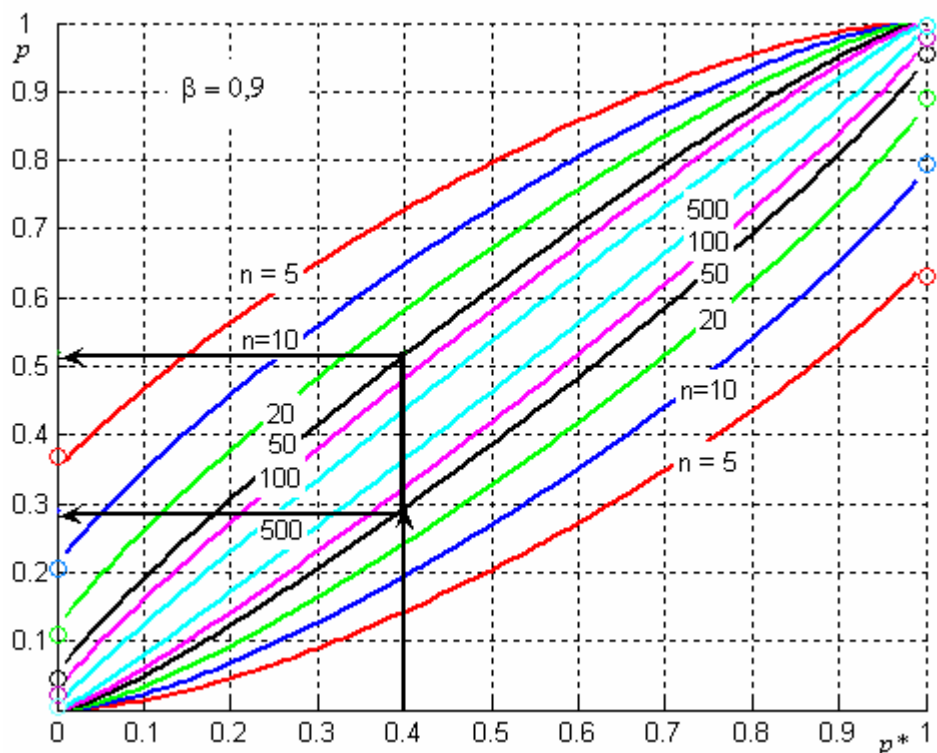


Рис. 2.3. Кривые для определения доверительных вероятностей по частоте событий

Оценка вероятности по частоте маловероятных событий

Кривые доверительных вероятностей на графике не доходят до нулевой и стопроцентной частоты, потому что построены с помощью нормального приближения, неприменимого в этих зонах. На осях $p^* = 0$ и $p^* = 1$ кружками соответствующих цветов отмечены точки, к которым должны прийти точно построенные кривые. При $p^* = 0$ эти точки означают верхние границы доверительного интервала для вероятностей ненаступления события (несрабатывания), если оно ни разу не наблюдалось в серии испытаний соответствующего объема (5, 10 и т.д.), а при $p^* = 1$ – надежность срабатывания. Если в n повторениях не наблюдалось ни одного отказа, нижняя граница доверительного интервала $p_1 = 0$, а верхнюю границу p_2 можно определить как наибольшую вероятность отказа, при которой полученный исход испытаний практически возможен, то есть имеет вероятность не меньше, чем $\alpha = 1 - \beta$:

$$p_{0,n} = (1 - p_2)^n > \alpha .$$

Из этого неравенства можно получить оценку надежности (отсутствия отказов) $1 - p_2 = \sqrt[n]{\alpha}$ и необходимое число испытаний $n = \ln(\alpha) / \ln(1 - p_2)$ для подтверждения данного уровня надежности с доверительной вероятностью β . Нижняя граница вероятности срабатывания после n успешных испытаний зависит от принятого уровня надежности следующим образом:

$$p_{n,n} = p_1^n > \alpha .$$

Указанные границы вычислены при $\alpha = 1 - \beta = 1 - 0,9 = 0,1$ и нанесены на график следующей командой:

```
>> N=[5 10 20 50 100 500]; plot(0, 1-(0.1).^(1./N),', 1, (0.1).^(1./N),'o')
```

При достаточно большом числе испытаний и малом числе успехов (или неудач) для определения границ доверительной вероятности можно воспользоваться пуассоновским приближением, в чем легко убедиться, заменив в

предыдущей команде $p_{0,n}$ и $p_{n,n}$ на p_0 и $1 - p_0$, выраженные по формуле Пуассона с параметрами, соответственно, np_2 и np_1 .

С помощью электронных формул можно разработать универсальный способ получения надежных оценок вероятности по частоте на всем интервале частостей, включая и ответственные участки в областях редких или почти достоверных успехов, где, как видно из рис. 2.3, приближенные формулы дают большие погрешности (кривые не ведут в центры кружков).

Определим нижнюю границу доверительного интервала p_1 как наименьшую вероятность успеха, при которой полученное в опыте число успехов еще возможно, то есть, превышает уровень $\alpha = 1 - \beta$, а верхнюю границу p_2 как наибольшую вероятность по такому же условию. К невероятным исходам опыта следует отнести наблюдаемое число успехов m или большее при определении нижней границы p_1 , и, аналогично, m или меньшее при определении верхней границы p_2 . Соответствующие суммы вероятностей p_k для $k \in [0, m]$ и $k \in [m, n]$ нужно вычислить с помощью электронной формулы `p_Binom`, чтобы избежать ошибок аппроксимации.

Именно так работает электронная формула `p_minmax` (Листинг 2.9) с теми же аргументами, что и `I_Binom`. Она дает более точные границы интервалов без завышения доверительной вероятности, причина которой объяснялась выше. Сравним ее результат с приведенным ранее примером (20 успехов в 50 опытах при доверительной вероятности 0,9):

```
>> p12= p_minmax(20,50,0.9)
p12 = 0.3058 0.5005
```

Этой формулой можно пользоваться и в предельных случаях:

```
>> p12= p_minmax(0,10,0.9), p34= p_minmax(10,10,0.9)
p12 = 0 0.2057
p34 = 0.7943 1.0000
```

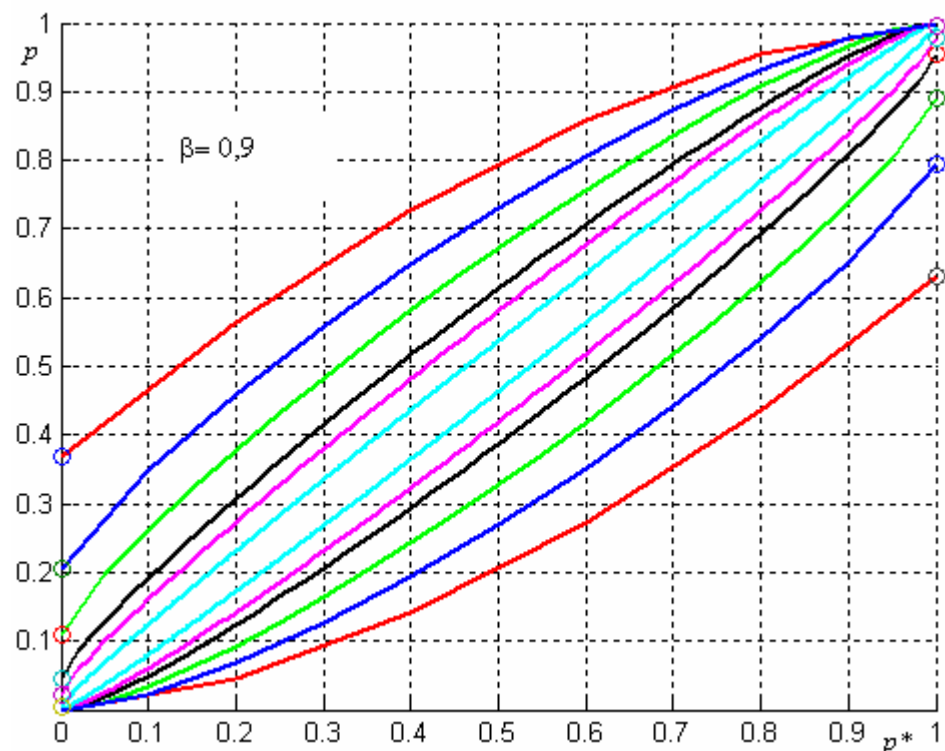


Рис. 2.3. Кривые для определения доверительных вероятностей по частоте событий

Более качественное, чем на рис. 2.3 семейство кривых (рис. 2.4) построено при n от 5 до 500 командами:

```
>> p1=[];p2=[];I=[0:n];for i=1 p12=p_minmax(i,n,P); p1(i+1)=p12(1);p2(i+1)=p12(2);end;
>> hold on,plot(I/n,[p2;p1], 'r')
```

Построение кривой чувствительности

С помощью электронной формулы p_{minmax} легко построить кривую вероятностей. Сначала генерируем статистику по известному закону:

```
>> n=15;x=1:n;p=1./(1+exp(-x+6)).^2;N=100;S=[];Y=[];for i=1:n S(i)=Gen('bin',N,p(i));end
```

По массиву частот срабатывания S вычислим нижние и верхние границы доверительного интервала для вероятностей срабатывания в каждой из n точек, в которых получены частоты:

```
>> for i=1:n Y(i,:)= p_minmax(S(i),N,0.99);end
```

Кривую чувствительности $F(x)$ получим как среднюю линию доверительного интервала:

```
>> F=mean(Y,2);plot(x,Y(:,1),x,Y(:,2),x,F,x,S/N,'k')
```

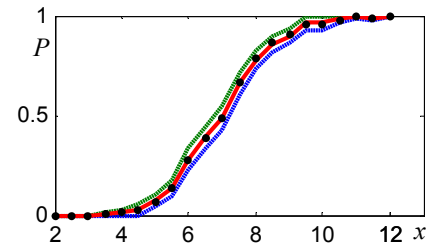


Рис. 2.4. Кривая чувствительности и доверительный интервал

На рис. 2.4 показаны графики границ доверительного интервала, кривой чувствительности и статистических точек. Графики довольно точно воспроизводят исходную кривую чувствительности благодаря очень большому объему статистики (15·500 точек). В физическом эксперименте объем получаемой статистики приходится ограничивать, что затрудняет выявление закономерностей.

Построение кривой чувствительности при ограниченном объеме статистики

По результатам эксперимента малого объема при $N=20$ испытаний в каждой из 10-и точек построим доверительные интервалы в каждой точке (рис. 2.5):

```
>> N=20;S=[];for i=1:n Y(i,:)= p_minmax(Gen('bin',N,p(i)),N,0.99);end
>> F=mean(Y,2);plot(x,Y(:,1),x,Y(:,2),x, F)
```

Кривая чувствительности должна плавно возрастать при увеличении аргумента, изломы объясняются случайными погрешностями из-за малого объема статистики. Электронная формула Approx (Приложение 26 Листинг 2.10) строит сглаживающую кривую по дискретным точкам, осуществляя параметрическую оптимизацию заданной зависимости по критерию минимума суммарной квадратичной погрешности:

```
>> [Fun,Par]=Approx('1./(1+exp(-x+L)).^2',x,F',1)
Fun(x,L) = 1./(1 +exp(-x+L)).^2 Par = 6.1404
```

Формула получает вид зависимости, массивы значений аргумента и аппроксимируемой функции, а также исходные значения для ее параметров. Она составляет целевую функцию оптимизации параметров для программы fminsearch и возвращает инлайн-функцию с оптимальными параметрами Par . Вычислим с помощью этой функции и выведем на рис. 2.5 (черным цветом) улучшенный график кривой чувствительности:

```
>> X=2:0.1:12; Y=Fun(X,Par);hold on, plot(X,Y, 'k')
```

Для сравнения выведена также кривая, полученная ранее из статистики большого объема (пунктиром). Они практически совпали, хотя улучшенная кривая получена по статистике в 25 раз меньшего объема.

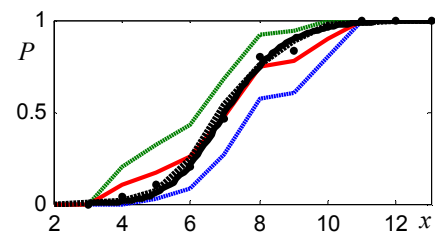


Рис. 2.5. Улучшенная кривая чувствительности при малом объеме статистики

Повторение опытов в меняющихся условиях

Вероятность m успехов в n независимых испытаниях, проводимых в *изменяющихся условиях* с вероятностью успеха p_i в i -м испытании, можно вычислить как коэффициент при z^m в разложении производящей функции

$$\varphi_n(z) = (p_1z + q_1)(p_2z + q_2)\dots(p_nz + q_n) \quad (2.11)$$

по степеням z :

$$\varphi_n(z) = p_n z^n + p_{n-1} z^{n-1} + \dots + p_0 z^0 \quad (2.12)$$

Универсальная электронная формула для независимых испытаний

Используя символьную операцию упрощения выражений `expand` для раскрытия скобок, можно создать процедуру вычисления коэффициентов в разложении (2.12), и заменить ею обращение к функции `p_Vinom`. Функция `RptTrial` может получить как одинаковую для всех повторений вероятность успеха, и тогда она вызывает `p_Vinom` с двумя аргументами, так и вектор вероятностей успеха в каждом повторении, тогда второй аргумент не нужен. В этой программе вычисление вероятностей $p_{m,n}$ готовит оператор

```
X=strcat(['',strep(char(expand(prod(P*Z+1-P))),'+',' '),']);
```

Он сначала составляет производящую функцию (2.11) от символьной переменной Z операцией умножения `prod`, преобразует ее в полином (2.12) функцией `expand`, затем превращает символьное выражение в строку символов алфавита (`char`), заменяет в ней символы '+' на пробелы (`strep`) и помещает в скобки как массив. Присваиванием символьной переменной X эта строка снова превращается в символьное выражение. Подстановка $Z=1$ в X дает массив вероятностей p_n, p_{n-1}, \dots, p_0 , который перестраивается в обратном порядке для совместимости со структурой результатов `Ber`.

Функцию `Ber` можно использовать с векторным или скалярными аргументами в виде `Ber([p1, ..., pn])` или `RptTrial(p, n)` как электронную формулу в любом контексте, где требуются вероятности $p_{m,n}, m = 0, \dots, n$.

Пример применения универсальной электронной формулы RptTrial

Пусть вероятность попадания в очереди уменьшается от 0,8 до 0,35, а закон поражения тот же, что и раньше $G(m) = 1 - 0,75^m$:

```
>> p=0.8:-0.05:0.35, n=length(p); G=1-0.75.^(0:n);
p = 0.80 0.75 0.70 0.65 0.60 0.55 0.50 0.45 0.40 0.35
```

Вычислим вероятности гипотез $p_{m,n}$ при переменных p_i , а также для трех постоянных значений $p_i = p_1 = 0,8, p_i = p_n = 0,35, p_i = p_{cp} = \sum p_i / n = 0,575$, построим для наглядности графики $p_{m,n}(m)$ и сравним все варианты вычисления гипотез по основному результату – полной вероятности:

```
>> P= Ber(p); P1= Ber(p(1),n); Pn= Ber(p(n),n); Pm=Ber(mean(p),n);
>> t=0:n; plot(t,P,t,P1,t,Pn,t,Pm), W=P*G', W1=P1*G', Wn=Pn*G', Wm=Pm*G'
W = 0.7900 W1 = 0.8926 Wn = 0.5998 Wm = 0.7882
```

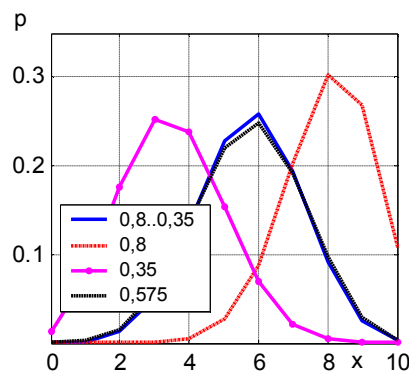


Рис. 2.2. Вероятности возможного числа попаданий

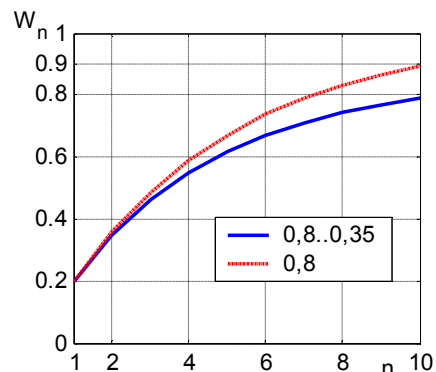


Рис. 2.3. Снижение вероятности поражения очереди с уменьшением вероятности попадания

Ординарные потоки и поля событий

Ординарные потоки событий

В последовательности испытаний Бернулли время несущественно, даже изменение вероятности успеха рассматривается не во времени, а по номеру испытания. Существует большой класс случайных явлений, в которых вероятность наступления того или иного события (сбой аппаратуры, поступление сигнала о цели) зависит, по крайней мере, от длительности интервала ожида-

ния события. Комплекс условий, определяющий случайные моменты времени наступления некоторого события, называется *поток событий*.

В *стационарном* потоке вероятность наступления события в данном интервале времени не зависит от момента начала интервала, а зависит только от его длительности. Это свойство упрощает анализ потока так же, как и неизменность условий в испытаниях Бернулли. Принципиальное значение, как и в испытаниях Бернулли, имеет независимость наступления событий. Так как на временной оси имеет смысл только независимость от прошлого, это свойство называют *отсутствием последействия*.

Еще одно фундаментальное свойство – *ординарность* – выделяет большой класс *редких событий* в том смысле, что два и более событий не могут наступить практически одновременно. Иначе говоря, в бесконечно малом промежутке времени вероятность наступления двух и более событий бесконечно мала по сравнению с вероятностью наступления одного события. Стационарный без последействия ординарный поток событий называется *простейшим пуассоновским потоком*.

Вероятность событий в простейшем пуассоновском потоке

В достаточно большом промежутке времени T число событий m статистически устойчиво, отношение $\lambda = m / T$ – среднее число событий в единицу времени – называется *плотностью потока*. В интервале $t \subset [0, T]$ число событий случайно с возможными значениями 0, 1, 2 и т.д. В стационарном потоке среднее число событий в любом интервале длительностью t составит $a = \lambda t$. Разделим весь период на элементарные интервалы Δt , достаточно малые, чтобы по условию ординарности в каждом из них могло бы наступить лишь одно событие. «Успешных» интервалов, в которых событие произошло, ровно m при общем их количестве $N = T / \Delta t$, вероятность выбрать случайно из всех интервалов «успешный» $p_1 = m / N = \lambda \Delta t$. Вероятность того, что в течение периода t наступит k событий, можно определить как число успехов в испытаниях Бернулли объемом $n = t / \Delta t$ и вероятностью успеха p_1 :

$$p_k = C_n^k p_1^k (1 - p_1)^{n-k}.$$

При $\Delta t \rightarrow 0$ выполняются условия теоремы Пуассона: $n \rightarrow \infty$, $p_1 \rightarrow 0$, $np_1 = \lambda t = a$. Следовательно, вероятность наступления k событий в интервале стационарного потока определяется по формуле Пуассона с параметром, равным произведению плотности потока на длительность интервала:

$$p_k = \frac{a^k}{k!} e^{-a}, k = 0, 1, 2, \dots \quad (2.13)$$

Пуассоновский поток событий

В нестационарном потоке среднее число событий в интервале $[t_0, t_0 + t]$ можно получить интегрированием переменной плотности по интервалу:

$$a = \int_{t_0}^{t_0+t} \lambda(\tau) d\tau, \quad (2.14)$$

после чего вероятность наступления k событий в интервале вычисляется по формуле (2.13). Нестационарные без последействия ординарные потоки событий называются *пуассоновскими*.

Простейшее пуассоновское поле событий

Поле событий называется область пространства, в каждой части которой может произойти случайное число событий (попаданий осколков в агрегат цели, наблюдаемых звезд на участке неба). Стохастически равномерное поле событий называется *однородным*. Среднее число событий в любой час-

ти однородного поля не зависит от формы выделенной части и расположения ее в поле, а пропорциональна мере этой части (объему, площади, длине). В *ординарном* поле вероятность наступления в бесконечно малой части поля двух и более событий пренебрежимо мала по сравнению с вероятностью наступления одного или ни одного события (бомбы в одно место дважды не попадают). Поле называют *независимым*, если на вероятность наступления в какой-то его части того или иного числа событий не влияет число событий в других частях поля.

Однородное, независимое, ординарное поле событий называется *простейшим пуассоновским полем*. В любой части такого поля единичной меры происходит в среднем одинаковое число событий λ (*плотность поля*), а в произвольной области поля D среднее число событий в $mes(D)$ раз больше: $a = \lambda mes(D)$. Условия независимости и ординарности позволяют утверждать, что вероятность наступления k событий в области D однородного поля с плотностью λ подчиняется закону Пуассона с параметром $a = \lambda mes(D)$.

Иллюстрация статистически равномерного распределения

Статистически равномерное распределение 500 точек с независимыми случайными координатами в квадрате 10×10 (рис. 2.4 а) формирует команда:

```
>> x=rand(1,500)*10;y=rand(1,500)*10;plot(x,y,''),grid on
```

Хотя никаких предпочтений при распределении точек нет, и на одну из 100 ячеек единичной площади приходится в среднем $\lambda = 500/10^2 = 5$ точек, случайное количество точек в ячейках может сильно отличаться от среднего. Подсчитаем число точек в каждой ячейке, чтобы установить частоту событий A_k (k точек в ячейке), $k = 0, 1, 2, \dots$:

```
>> N=[];I=0:1:10;n=length(I)-1;
>> for i=1:n for j=1:n N(i,j)=sum(I(i)<x & x<I(i+1) & I(j)<y & y<I(j+1));end,end
>> for j=0:12 A(j+1)=sum(sum(N==j)); end, A
A = 1 5 8 11 19 18 13 12 6 4 1 1 0
```

Результат в массиве A показывает, что одна ячейка пустая, пять ячеек содержат по одной точке, 8 ячеек – по две и т.д. Значения элементов массива A случайны, их отношение к общему числу ячеек – это частота p_k^* событий A_k . Графики частот $p_k^*(k)$ и вероятностей тех же событий $p_k(k)$, вычисленных по формуле Пуассона (рис. 2.4, б), построим командой:

```
>> plot(t,p_Poisson(5,t),t,A(t+1)/n^2)
```

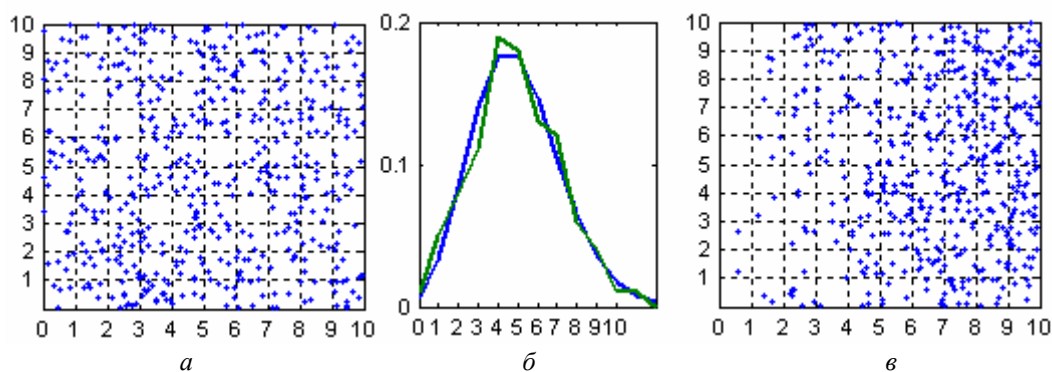


Рис. 2.4. Однородное (а) и неоднородное (б) пуассоновские поля

Близость эмпирических частот и вероятностей подтверждает применимость формулы Пуассона для вычисления вероятности событий A_k . Убедимся в ординарности поля, разделив его на еще более мелкие части:

```
>> N=[];I=0:0.1:10;n=length(I)-1;
>> for i=1:n for j=1:n N(i,j)=sum(I(i)<x&x<I(i+1)&I(j)<y&y<I(j+1));end,end
>> for j=0:12 A(j+1)=sum(sum(N==j)); end, A
A = 9511 478 11 0 0 0 0 0 0 0 0 0
```

Из 10000 ячеек (не бесконечно малых) только в 11 попало две точки, остальные – либо пусты, либо содержат по одной точке.

Пуассоновское поле событий

В пуассоновском поле (не простейшем) выполняются условия независимости и ординарности, но плотность не одинакова в различных частях поля. Те же 500 точек, что и в предыдущем примере, распределены на рис. 2.4 *в* с плотностью $\lambda(x, y) = x$. Это поле сформировано командой, которая обеспечивает равномерное распределение в интервале $[0, 1]$ ординат параболы на отрезке $[0, 10]$ (теоретическое обоснование этой процедуры будет дано в Лекции 9):

```
>> x=sqrt(rand(1,500))*10;y=rand(1,500)*10;plot(x,y,'*'),grid on
```

Среднее число точек поля в вертикальных полосах единичной ширины должно быть примерно равным абсциссе середины полосы (0,5, 1,5 и т.д.):

```
>> N=[];I=0:1:10;n=length(I)-1; for i=1:n N(i)=sum(I(i)<x&x<I(i+1));end, A=N/10  
A = 0.40 1.70 2.10 3.40 4.50 5.10 6.60 7.00 8.80 9.40
```

Среднее число событий в произвольной области D пуассоновского поля получается интегрированием плотности по этой области

$$a = \int_D \lambda(s) ds, \quad (2.15)$$

где ds – бесконечно малая часть области D . Численное интегрирование не обязательно. Учитывая дискретность поля событий, его плотность считают постоянной на малых, но конечных участках (проекциях УА). Тогда среднее число событий на i -й части, мера которой S_i , составляет $a_i = \lambda_i S_i$.

Программа верификации кода MATLAB

```

clear all
p=0.75;eps=0.1; p1=p-eps;p2=p+eps; Pk=0.9;
for n=9:90 m=ceil(n*p1):fix(n*p2); P=sum(Ber(p,n,m));if P>Pk break,end,end,n,P
x=0:0.1:0.9;X=[x' 1+x' 2+x'];f=f_Gauss(X);Fi=F_LaplasV(X);
disp([X(:,1) f(:,1) Fi(:,1) X(:,2) f(:,2) Fi(:,2) X(:,3) f(:,3) Fi(:,3)])
x=ArgLaplas(0.9/2)
n=3/4*1/4*(ArgLaplas(0.9/2)/0.1)^2;n=round(n),m12=round(n*[0.65 0.85])
P=sum(Ber(p,51,33:43))
K=0.92;for n=10:100 a=ceil(n*p1):fix(n*p2);P=sum(Ber(p,n,a));if P>K
break,end,end,n,P
n=1/2*1/2*(ArgLaplas(0.9/2)/0.1)^2
R4=sum(p_Poisson(1,0:3))

clear all
R_4=sum(p_Binom(0.01,100,0:3))
n=10;p=0.1;R4=sum(p_Poisson(p*n,0:3)),R_4=sum(p_Binom(p,n,0:3))

clear all
P=0.9;T=ArgLaplas(P/2)^2;x=0.01:0.01:0.99; C=x.^2; hold on
n=5; A=1+T/n;B=x+T/n/2;D=sqrt(B.^2-A*C);p2=(B+D)/A;p1=(B-D)/A;plot(x,[p2;p1])
n=50;x=0.4;C=x^2;t=T/n;A=1+t;B=x+t/2;D=sqrt(B.^2-A*C);p2=(B+D)/A;p1=(B-D)/A
N=[5 10 20 50 100 500]; plot(0, 1-(0.1).^(1./N),'.', 1, (0.1).^(1./N),'o')
p12= p_minmax(20,50,0.9)
p12= p_minmax(0,10,0.9), p34= p_minmax(10,10,0.9)
p1=[];p2=[];I=[0:n];for i=I p12=p_minmax(i,n,P);
p1(i+1)=p12(1);p2(i+1)=p12(2);end;
hold on,plot(I/n,[p2;p1],'r')

clear all, clear figure
n=15;x=1:n;p=1./(1+exp(-x+6)).^2;N=100;S=[];Y=[];for i=1:n
S(i)=Gen('bin',N,p(i));end
for i=1:n Y(i,:)= p_minmax(S(i),N,0.99);end
F=mean(Y,2);plot(x,Y(:,1),x,Y(:,2),x,F,x,S/N,'k. ')

clear figure
N=20;S=[];for i=1:n Y(i,:)= p_minmax(Gen('bin',N,p(i)),N,0.99);end
F=mean(Y,2);plot(x,Y(:,1),x,Y(:,2),x,F)
[Fun,Par]=Approx('1./(1+exp(-x+L)).^2',x,F',1)
X=2:0.1:12; Y=Fun(X,Par);hold on, plot(X,Y, 'k')

clear all
p=0.8:-0.05:0.35, n=length(p); G=1-0.75.^(0:n);
P= Ber(p); P1= Ber(p(1),n); Pn= Ber(p(n),n); Pm=Ber(mean(p),n);
t=0:n; plot(t,P,t,P1,t,Pn,t,Pm), W=P*G', W1=P1*G', Wn=Pn*G', Wm=Pm*G'

clear all
x=rand(1,50)*10;y=rand(1,50)*10;plot(x,y, '.'),grid on
N=[];I=0:1:10;n=length(I)-1;
for i=1:n for j=1:n N(i,j)=sum(I(i)<x & x<I(i+1) & I(j)<y & y<I(j+1));end,end
for j=0:12 A(j+1)=sum(sum(N==j)); end, A
N=[];I=0:0.1:10;n=length(I)-1;
for i=1:n for j=1:n N(i,j)=sum(I(i)<x&x<I(i+1)&I(j)<y&y<I(j+1));end,end
for j=0:12 A(j+1)=sum(sum(N==j)); end, A

clear all
x=sqrt(rand(1,50))*10;y=rand(1,50)*10;plot(x,y, '*'),grid on
N=[];I=0:1:10;n=length(I)-1; for i=1:n N(i)=sum(I(i)<x&x<I(i+1));end, A=N/10

clear all
n=40; p=0.1; B=Ber(p,n); P=p_Poisson(n*p,n);
x=0:0.1:n; s=sqrt(n*p*(1-p)); G=f_Gauss((x-n*p)/s)/s;
plot(0:n, P, 0:n, B, x, G)

```

Контрольные вопросы и задачи

1. Чем отличается стохастически равномерное поле от равномерного распределения узловых точек в узлах равномерной сетки?
2. Случайное событие $A_{2,4}$ – ровно $m = 2$ попадания в $n = 2$ независимых выстрелах. Выразите это событие через события A_k – попадание в k -м выстреле. Примените к полученному выражению теоремы сложения и умножения вероятностей. Обобщите результат для произвольных m, n .
3. Какова вероятность не менее $K = 3$ успехов в $n = 10$ испытаниях Бернулли с вероятностью успеха в каждом повторении $p = 0,3$?
4. Как вычислить вероятность не более $K = 3$ успехов в $n = 100$ испытаниях Бернулли с вероятностью успеха в каждом повторении $p = 0,03$?
5. Как получено выражение для вероятности данной комбинации числа событий в серии испытаний Бернулли с более, чем двумя возможными исходами?
6. Как получено выражение для вероятности данного числа успехов в серии независимых испытаний, когда вероятности успехов в разных испытаниях не одинаковы?
7. Объясните доказательство теоремы Пуассона.
8. На каком основании можно считать осколочное поле пуассоновским или простейшим пуассоновским?
9. Простейшее пуассоновское поле точек на плоскости имеет плотность 0,5. Какова вероятность попадания хотя бы одной точки в круг единичного радиуса?
10. Плотность нестационарного пуассоновского потока отказов линейно возрастает по закону $\lambda(t) = a + b t$. Какая вероятность того, что в интервале времени $[t_1, t_2]$ не будет отказов?
11. Объясните вероятностный смысл координатного закона поражения.
12. Объясните вероятностный смысл функции уязвимости.
13. Чем объясняется возрастание КЗП от начала координат на рис. 2.8?
14. Выполнить количественную оценку степени приближения легко с помощью электронных формул. Построить многоугольники биномиального распределения при $n = 40, p = 0,1$ и Пуассона с параметром $\lambda = np = 4$, а также график аппроксимирующей функции (2.9), повторить эту процедуру при $n = 40, p = 0,2$ и $n = 40, p = 0,05$.

Решение:

```
>> n=40; p=0.1; B=Ber(p,n); P=p_Poisson(n*p,n);
>> x=0:0.1:n; s=sqrt(n*p*(1-p)); G=f_Gauss((x-n*p)/s)/s;
>> plot(0:n, P, 0:n, B, x, G)
```

Сравнение отклонений графиков нормального и пуассоновского приближений от результатов расчета по биномиальной формуле на рис. 2.1 показывает, в какой степени погрешности нормального приближения растут с приближением p к границам интервала $[0, 1]$, а пуассоновского – при увеличении произведения np (положения максимума биномиального распределения) от 2 до 8. В варианте $p = 0,05$ и $np = 2$ точнее пуассоновское приближение, а $p = 0,2$ и $np = 8$ – нормальное. Это нужно учитывать, выбирая корректное приближение биномиальной формулы для решения практических задач.

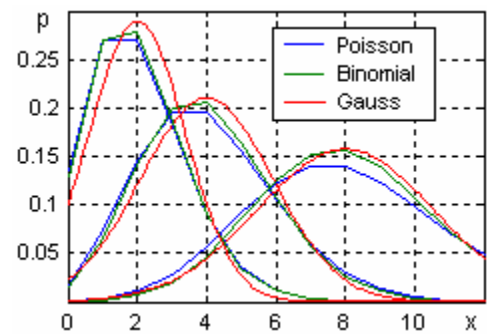


Рис. 2.1. Погрешности аппроксимации биномиальной формулы