Предисловие редактора перевода

Издательство «БИНОМ. Лаборатория знаний» продолжает выпуск серии монографий по теории вероятностных распределений. Серия является переводом многотомного англоязычного издания, написанного Н. Джонсоном, С. Коцем, Н. Балакришнаном, А. У. Кемп, которое было выпущено издательством Wiley, и состоит из шести томов. Том 1 посвящен одномерным дискретным распределениям, тома 2 и 3 — одномерным непрерывным распределениям, том 4 — многомерным дискретным, а тома 5 и 6 — многомерным непрерывным распределениям.

Такая классификация распределений позволила авторам в тт. 1, 2, 4 и 5 изложить необходимый общий математический аппарат, используемый при анализе свойств конкретных типов распределений и многочисленных классов семейств распределений.

В каждой книге серии приводятся: аналитические свойства семейств распределений, формулы и алгоритмы, необходимые для статистического анализа данных и компьютерной генерации соответствующих псевдослучайных величин, обширные сведения и библиография по исследованию свойств и приложениям соответствующих семейств распределений.

По структуре, широте и глубине охвата материала данная серия является, по существу, энциклопедическим справочником, предназначенным для научных работников и исследователей-прикладников естественной, гуманитарной и экономической сфер.

Для студентов и преподавателей — это своеобразное учебное пособие по теории вероятностей, математической статистике и их приложениям.

Весьма полная аннотированная библиография по исследованию свойств различных распределений позволяет рассматривать данное издание в качестве своеобразного путеводителя при построении статистических моделей реальных явлений.

Во втором томе гл. 12–13 переведены Н. А. Шиховой, гл. 15 — О. И. Волковой, гл. 14, 16–21 — М. С. Стригуновой.

Е. В. Чепурин

Посвящается:
Реджине Эландт—Джонсон, Розали Коц, Колин Катлер и Саре Балакришнан

Предисловие

В качестве продолжения второго издания книги *Одномерные дискретные распределения* настоящая монография является первым из двух томов, посвященных непрерывным одномерным распределениям. Второе издание книги *Одномерные непрерывные распределения* отличается от первого, вышедшего в свет в 1970 г., в следующих двух важных аспектах. (1) К коллективу авторов присоединился профессор Н. Балакришнан. (2) В силу существенных продвижений в теории, методологии и практике применения непрерывных распределений, в особенности, гамма-, Вейбулла и обратного гауссовского распределений, за последние 20 лет, было решено перенести главу, посвященную распределению экстремальных значений, в следующий том. Глава, посвященная гамма-распределению, была разбита на две: одна из них целиком посвящена распределению хи-квадрат. Даже при этих условиях, так же, как и при пересмотре тома *Одномерные дискретные распределения*, большое количество дополнительной информации, накопившейся со времени выпуска первого издания, привело к существенному увеличению объема.

В соответствии с принципом, сформулированным в *Предисловии к серии монографий*, мы продолжаем преследовать цель «исключения теоретических изысканий, не имеющих видимой практической значимости», хотя мы включаем материал о характеризациях, который можно считать имеющим сомнительное практическое применение. Глава 12, носящая общий характер, претерпела меньшие изменения по сравнению с другими главами, посвященными конкретным законам.

Даже после исключения главы, посвященной распределению экстремальных значений, большое количество новой поступившей информации вынудило нас быть очень избирательными по отношению к новым работам. В частности, мы стремились бороться с разрозненностью фактов, необходимость чего была изящно выражена проф. А. П. Дэвидом, новым редактором журнала *Biometrika* (см. [Biometrika, **80**, 1 (1993)]). Мы прекрасно понимаем, что некоторые авторы могут быть обижены отсутствием упоминания их работ, но мы надеемся, что они расценят это не как недоброжелательность, а «в лучшем случае» как следствие нашей невежественности. Настоящая серия монографий планировалась скорее для пользы читателей, чем как «почетный список» научных работ.

За мастерски выполненный набор текста мы благодарим миссис Лизу Брукс (Университет Северной Каролины), миссис Синди Паттерсон (Государственный Университет Боулинг Грин) и миссис Дебби Иско (Гамильтон, Канада). Мы также выражаем признательность работникам библиотек Университета Северной Каролины, Государственного Университета Боулинг Грин, Университета МакМастер, Университета Ватерлоо, а также Университета Мэрилэнда за их помощь при поиске литературы. Работа над настоящей монографией была выполнена: Самюэлем Коцем в качестве почетного приглашенного профессора факультета Математики и Статистики Государственного Университета Боулинг Грин (Боулинг Грин, шт. Огайо) в период с сентября по декабрь 1992 г., проф. Н. Балакришнаном в период творческого отпуска с июля 1992 г. по июнь 1993 г. на факультете Статистики и Актуарных дисциплин Университета Ватерлоо.

Особая наша благодарность миссис Кэйт Роач и мистеру Эду Кантиллону из нью-йоркского отделения издательства John Wiley & Sons за их искренние усилия по обеспечению высокого качества настоящего издания. Мы также выражаем признательность мисс Дане Эндрюс за редактирование рукописи.

Мы очень благодарны следующим организациям: Institute of Mathematical Statistics, the American Statistical Association, the Biometrika Trustees, the Institute of Electrical and Electronic Engineerings, the Association for Computing Machinery, Marcel Dekker, Inc., the Australian Statistical Society, Gordon and Breach Science Publishers, Blackwell Publishers, а также редакторам журналов *Biometrical Journal*, *Sankhyā* и *Tamkang Journal of Mathematics* за разрешение использования ранее опубликованных таблиц и рисунков.

При написании этого, в некотором роде, информационного обзора — расчитанного в первую очередь на неспециалистов — авторы неожиданно столкнулись с необходимостью пояснения результатов, возможно очевидных для специалистов, но не являющихся общеизвестными. Такие результаты несут в себе много информации, которая может оказаться новой и ценной для знатоков. Существует опасность, с одной стороны, «сильной популяризации» (в противоположность упрощению) результатов, а, с другой стороны, чрезмерного акцента на плохо изученных вопросах — в котором многие читатели не нуждаются. Мы пытались насколько возможно избежать этих ошибок.

Мы искренне надеемся, что предлагаемая серия монографий обеспечит «заинтересованного читателя» полезными для его работы результатами, а также напомнит эти результаты читателям искушенным, побуждая их к самостоятельному анализу и дальнейшим научным исследованиям.

Н. Л. Джонсон (*N. L. Johnson*),
С. Коц (*S. Kotz*),
А. У. Кемп (*A. W. Kemp*)

Глава 12

Непрерывные распределения (общие сведения)

1. Введение

Главы 1 и 2 (тома, посвященного дискретным распределениям[1]) содержат ряд общих результатов и методов, применимых и к непрерывным распределениям. В настоящей главе мы пополним их сведениями, относящимися непосредственно к непрерывным распределениям. И так же, как в гл. 2 (тома «Одномерные дискретные распределения»), опишем несколько общих *семейств* (непрерывных) распределений.

Обычно непрерывные распределения поддаются анализу, математически более элегантному, чем дискретные. Это делает их особенно полезными в качестве приближений дискретных распределений. Такое применение непрерывных распределений используется во многих приложениях как при построении моделей, так и при использовании статистических методов. Непрерывные распределения использовались при аппроксимации дискретных распределений статистик в томе «Одномерные дискретные распределения». Тот факт, что большинство применений непрерывных распределений в построении моделей сводятся к аппроксимации дискретных распределений, возможно, не столь высоко ценится, но является не менее верным. Очень редко представляется более разумным, в некотором абсолютном смысле, представить наблюдаемую переменную не дискретной, а непрерывной случайной величиной. Такое представление является, скорее, удобным *приближением*, облегчающим математическую и статистическую обработку.

Отличительным свойством непрерывной случайной величины является то, что любое фиксированное значение из пространства своих значений она принимает с нулевой вероятностью. В общем случае ненулевой является вероятность, вычисленная как определенный интеграл от *плотности распределения вероятностей* (разд. 1.4) того, что она примет значение из фиксированного (конечного или бесконечного) интервала. Когда наблюдаемая переменная описывается непрерывной случайной величиной, ее регистрируемые значения по необходимости «дискретизированы». Например, если измерения проводятся с точностью 0.01, то все значения из интервала $(8.665, 8.675)$ будут записаны числами 8.67. Тем самым данные оказываются *сгруппированными*. Поправки

[1] Здесь и далее имеется в виду монография *Джонсон Н. Л., Коц С., Кемп А. У.* Одномерные дискретные распределения. — М. : БИНОМ, Лаборатория знаний, 2009. В ней представлены главы с 1-й по 11-ю. — *Прим. ред.*

в оценивании моментов генеральной совокупности для устранения (в среднем) этого эффекта были предложены Шеппардом (см. [Sheppard (1896)]). Выпишем их для случая, когда группировка производится посредством интервалов одинакового размера h (далее через μ'_r и ${}_g\mu'_r$ обозначены, соответственно, r-е моменты до и после группировки):

$$\mu'_1 = {}_g\mu'_1, \tag{12.1a}$$

$$\mu'_2 = {}_g\mu'_2 - \frac{1}{12}h^2, \tag{12.1b}$$

$$\mu'_3 = {}_g\mu'_3 - \frac{1}{4}\,{}_g\mu'_3 h^3, \tag{12.1c}$$

$$\mu'_4 = {}_g\mu'_4 - \frac{1}{2}\,{}_g\mu'_2 h^2 + \frac{7}{240}h^4. \tag{12.1d}$$

Общая формула имеет вид (см. [Sheppard (1896)], [Wold (1934)])

$$\mu'_r = \sum_{j=0}^{r} \left(2^{1-j} - 1\right) \binom{r}{j} B_j \cdot {}_g\mu'_{r-j}\, h^j, \tag{12.2}$$

где B_j — j-е число Бернулли (см. гл. 1, разд. A9). Эти формулы выписаны в предположении, что центры групп находятся в точках

$$\ldots,\ a-h,\ a,\ a+h,\ \ldots \qquad \left(|a| < \frac{1}{2}\,h\right),$$

а точка a имеет равномерное распределение (см. гл. 26) на отрезке $\left[-\frac{1}{2}h;\ \frac{1}{2}h\right]$ (см. также [Haitovsky (1983)]). Влияние необоснованности этого предположения исследовано в работе [Tricker (1984)], где показано, что характеристическая функция (гл. 1, разд. B8) случайной величины ${}_gX$ («сгруппированной» величины, отвечающей случайной переменной X) имеет вид

$$\varphi_{gX}(t) = \sum_{j=-\infty}^{\infty} \frac{\sin\left(\frac{1}{2}\,\mathrm{th} + j\pi\right)}{\frac{1}{2}\,\mathrm{th} + j\pi} \exp\left(1 - i\cdot 2\pi jh\right)\cdot \varphi_X\left(\mathrm{th} + 2\pi jh\right), \tag{12.3}$$

где $i = \sqrt{-1}$, а $\varphi(t)$ — характеристическая функция случайной величины X.

В той же статье численно оценено влияние округления в случаях, когда переменная X имеет нормальное распределение (гл. 13), распределение Лапласа (гл. 24) и гамма-распределение (гл. 17). При заданной ширине группирования h абсолютная величина поправки к группировке увеличивается при росте абсолютного значения асимметрии $|\sqrt{\beta_1}|$; она чрезвычайно мала для симметричных распределений.

Некоторые понятия, имеющие большую значимость для дискретных распределений, гораздо менее важны в случае непрерывных распределений. В частности, производящая функция вероятностей мало используется в этой части книги. Факториальные моменты также редко обладают преимуществами краткости и простоты, которые присутствовали в случае дискретных распределений, хотя эти моменты и можно вычислить.

В случае непрерывных распределений гораздо более полезно *нормирование*, т. е. использование случайной величины

$$\frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}(X)}},$$

имеющей распределение с нулевым средним и единичным стандартным отклонением. В частности, форму распределения удобно описать заданием нормированных значений нескольких *квантилей* (т. е. значений случайной величины, в которых функция распределения принимает заданные значения). Следует различать *нормированную* и *стандартную* формы распределения. Последняя обычно удобна для получения формул, связанных с функцией плотности. Она *может* совпасть с нормированной формой, но это необязательно.

В работе [MacGillivray (1992)] вводится *функция асимметричности*

$$\gamma_X(u) = \frac{F^{-1}(u) + F^{-1}(1-u) - 2F^{-1}\left(\frac{1}{2}\right)}{F^{-1}(u) - F^{-1}(1-u)}. \tag{12.4a}$$

Заметим, что $F^{-1}\left(\frac{1}{2}\right) = \text{медиана}\,(X)$ и что $\gamma_X\left(\frac{3}{4}\right)$ является мерой асимметрии Гальтона:

$$\frac{(\text{верхняя квартиль} - \text{медиана}) - (\text{медиана} - \text{нижняя квартиль})}{\text{межквартильное расстояние}}.$$

Величина

$$\sup_{\frac{1}{2}\leqslant u<1} |\gamma_X(u)| \tag{12.4b}$$

предлагается (см. [MacGillivray (1992)]) в качестве «меры общей асимметрии для центральной $100(1-2\alpha)$%-й части распределения».

Дадим определения коэффициентов, наиболее часто употребляемых для характеристики формы именно в случае непрерывных распределений.

1. *Усредненная разность Джини* $\gamma(X)$ для распределения случайной переменной X равна математическому ожиданию $\mathrm{E}[|X_1 - X_2|]$ абсолютного значения разности между двумя независимыми случайными величинами X_1 и X_2, распределение каждой из которых совпадает с распределением случайной переменной X.

Если $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины, то статистика

$$g = \binom{n}{2}^{-1} \sum_{i<j}^{n}\sum^{n} |X_i - X_j| \tag{12.5}$$

является несмещенной оценкой величины

$$\gamma(X) = \mathrm{E}[|X_1 - X_2|]. \tag{12.6}$$

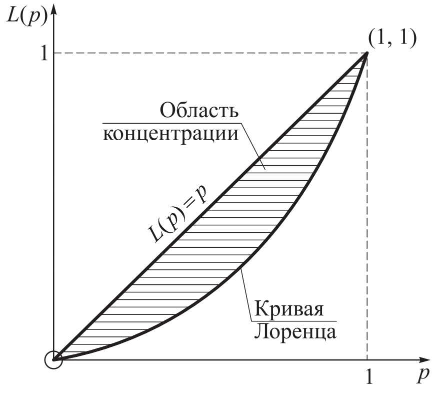

Рис. 12.1. Кривая Лоренца

2. *Кривая Лоренца* для положительной случайной величины X является графиком отношения
$$L\left(F_X(x)\right) = \frac{\mathrm{E}\left[X \mid X \leqslant x\right] F_X(x)}{\mathrm{E}\left[X\right]} \tag{12.7}$$
в зависимости от значений $F_X(x)$. Если случайная переменная X представляет годовой доход, то величина $L(p)$ есть доля общего дохода, полученного индивидуумами, имеющими по крайней мере $100p$%-й доход.

Нетрудно видеть, что
$$L(p) \leqslant p,$$
$$L(0) = 0,$$
$$L(1) = 1.$$

Типичная кривая Лоренца показана на рис. 12.1. Если доход всех индивидуумов равномерен, то $L(p) = p$. Площадь фигуры, ограниченной прямой $L(p) = p$ и кривой Лоренца, можно рассматривать как меру неравномерности дохода или, в более общем виде, как меру изменчивости распределения случайной величины X. В работе [Gail and Gastwirth (1978)] подробно рассматриваются кривые Лоренца, а краткое изложение их свойств можно найти в статье [Dagum (1985)].

В статье [Gastwirth (1971)] приводится следующее определение функции Лоренца $L(p)$:
$$L(p) = \left\{\mathrm{E}\left[X\right]\right\}^{-1} \int\limits_0^p F_X^{-1}(t)\, dt, \tag{12.8}$$
где
$$F_X^{-1}(t) = \inf_x \left\{x : F_X(t) \geqslant t\right\}.$$

Это определение применимо и в случае дискретных распределений, а для непрерывных распределений равносильно соотношению (12.7).

Для сравнения степеней изменчивости двух или более распределений иногда используют частичное *упорядочение по Лоренцу*. Оно основано на сравнении значений функций $L(p)$ для этих распределений. Если разности этих значений имеют один и тот же знак при всех p, то распределения можно упорядочить по Лоренцу. В противном случае упорядочение по Лоренцу неприменимо. В книге [Arnold (1987)] приводится полезное описание такого упорядочения; см. также гл. 33.

Фигура, ограниченная прямой $L(p) = p$ и кривой Лоренца, называется *областью концентрации*. *Индекс концентрации Джини* $C(X)$ равен удвоенной площади этой фигуры. Покажем его связь с *усредненной разностью Джини*, определенной формулой (12.6):

$$C(X) = 2\int_0^1 \{p - L(p)\}\, dp = 1 - 2\int_0^p L(p)\, dp.$$

Далее,

$$\begin{aligned}\int_0^p L(p)\, dp &= \frac{1}{\mathrm{E}[X]}\int_0^1 \mathrm{E}[X \mid X < x]\; F_X(x)\, dF_X(x) = \\ &= \frac{1}{\mathrm{E}[X]}\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{x} tp_X(t)\, dt\right\} p_X(x)\, dx = \\ &= \frac{1}{\mathrm{E}[X]}\,\mathrm{E}[X_1 \mid X_1 < X_2]\;\Pr[X_1 < X_2],\end{aligned}$$

где X_1 и X_2 — независимые случайные величины, распределение каждой из которых совпадает с распределением случайной переменной X. Так как X_1 и X_2 непрерывны и независимы, то

$$\Pr[X_1 < X_2] = \Pr[X_1 > X_2] = \frac{1}{2};$$

отсюда следует, что

$$C(X) = 1 - \frac{\mathrm{E}[X_1 \mid X_1 < X_2]}{\mathrm{E}[X]}.$$

Так как

$$\frac{1}{2}\{\mathrm{E}[X_1 \mid X_1 < X_2] + \mathrm{E}[X_1 \mid X_1 > X_2]\} = \mathrm{E}[X]$$

и

$$\gamma(X) = \mathrm{E}[|X_1 - X_2|] = \mathrm{E}[X_1 \mid X_1 > X_2] - \mathrm{E}[X_1 \mid X_1 < X_2],$$

то

$$\mathrm{E}[X_1 \mid X_1 < X_2] = \mathrm{E}[X] - \frac{1}{2}\gamma(X).$$

Следовательно,

$$C(X) = \frac{\frac{1}{2}\gamma(X)}{\mathrm{E}[X]}. \qquad (12.9\text{a})$$

Отношение

$$\frac{\gamma(X)}{\mathrm{E}[X]} = \frac{\mathrm{E}[|X_1 - X_2|]}{\mathrm{E}[X]}$$

аналогично коэффициенту изменчивости. В работе [Tziafetas (1989)] получено другое выражение для коэффициента Джини:

$$C(X) = \frac{\mathrm{Cov}(X,\ F_X(X))}{\mathrm{E}[X]}\ . \tag{12.9b}$$

В случае непрерывных распределений порядковые статистики используются гораздо чаще, чем в случае дискретных; и они проще для теоретического изучения. Следующий раздел посвящен общим сведениям о порядковых статистиках непрерывных случайных величин с особым акцентом на их использование в статистическом анализе.

2. Порядковые статистики

Пусть X_1, $X_2, \ldots, X_n$ — случайные величины. Обозначим через $X_1' \leqslant X_2' \leqslant \ldots \leqslant X_n'$ те же переменные, упорядоченные по неубыванию (так, что $X_1' = \min(X_1, X_2, \ldots, X_n)$, $X_n' = \max(X_1, X_2, \ldots, X_n)$). Тогда X_1', $X_2', \ldots, X_n'$ называются *порядковыми статистиками*, соответствующими случайным величинам X_1, $X_2, \ldots, X_n$ (см. также гл. 1, разд. B10). При необходимости явно указать количество случайных величин, для которых определяются соответствующие порядковые статистики, используются обозначения $X_{1:n}'$, $X_{2:n}', \ldots, X_{n:n}'$.

Если разности $\{X_i - X_j\}$ суть непрерывные случайные величины, то все события $\{X_i = X_j\}$ имеют нулевую вероятность. Так как число их конечно, при вероятностных расчетах ими можно пренебречь. Далее мы будем считать, что $X_1' < X_2' < \ldots < X_n'$, что никак не отразится на вероятностях, связанных с совместным распределением статистик X_1', $X_2', \ldots, X_n'$.

Функция распределения статистики X_n' определяется равенством

$$\Pr\left[X_n' \leqslant x\right] = \Pr\left[\bigcap_{j=1}^{n} \left(X_j \leqslant x\right)\right]. \tag{12.10}$$

Если случайные величины X_1, $X_2, \ldots, X_n$ взаимно независимы, то

$$\Pr\left[X_n' \leqslant x\right] = \prod_{j=1}^{n} \Pr\left[X_j \leqslant x\right], \tag{12.11}$$

и плотность распределения статистики X'_n равна[1)]

$$\begin{pmatrix}\text{функция распределения}\\ \text{статистики } X'_n\end{pmatrix} \times \sum_{j=1}^{n} \left(\frac{\begin{matrix}\text{плотность распределения}\\ \text{величины } X_j\end{matrix}}{\begin{matrix}\text{функция распределения}\\ \text{величины } X_j\end{matrix}} \right). \tag{12.12}$$

Если все случайные величины X_j одинаково распределены и $\Pr\left[X_j \leqslant x\right] = F(x)$, $dF(x)/dx = p(x)$, то плотность распределения статистики X'_n равна

$$n\,[F(x)]^{n-1}\,p(x). \tag{12.13a}$$

Аналогичным образом выписывается выражение для плотности распределения статистики X'_1 (снова в предположении независимости и одинаковой распределенности случайных величин X'_1):

$$n\,[1 - F(x)]^{n-1}\,p(x). \tag{12.13b}$$

В общем случае для индексов $1 \leqslant a_1 < a_2 < \ldots < a_s \leqslant n$ (считается, что $a_0 = 0$, $a_{s+1} = n$, $F\left(x_{a_0}\right) = 0$, $F\left(x_{a_{s+1}}\right) = 1$) плотность совместного распределения статистик $X'_{a_1}, X'_{a_2}, \ldots, X'_{a_s}$ равна

$$\frac{n!}{\prod_{j=1}^{s+1}\left(a_j - a_{j-1} - 1\right)!}\left[\prod_{j=1}^{s+1}\left\{F\left(x_{a_j}\right) - F\left(x_{a_{j-1}}\right)\right\}^{a_j - a_{j-1} - 1}\right]\prod_{j=1}^{s} p\left(x_{a_j}\right)$$
$$\left(x_{a_1} \leqslant x_{a_2} \leqslant \ldots \leqslant x_{a_s}\right). \tag{12.14}$$

В частности, плотность совместного распределения экстремальных статистик X'_1 и X'_n имеет вид

$$n(n-1)p\left(x_1\right)p\left(x_n\right)\left[F\left(x_n\right) - F\left(x_1\right)\right]^{n-2} \qquad \left(x_1 \leqslant x_n\right). \tag{12.15}$$

Используя выражение для этого совместного распределения, можно получить следующую функцию распределения размаха $W = X'_n - X'_1$:

$$\Pr[W \leqslant w] = n \int_{-\infty}^{\infty} p(x)\,[F(x) - F(x-w)]^{n-1}\,dx. \tag{12.16}$$

При этом

$$\mathrm{E}[W] = \int_{-\infty}^{\infty}\left\{1 - [F(x)]^n - [1 - F(x)]^n\right\}dx. \tag{12.17}$$

Если число $n = 2m + 1$ нечетно (т. е. m — целое), то статистика X'_{m+1} называется *выборочной медианой* в выборке $X_1, X_2, \ldots, X_n$. Плотность ее распределения равна

$$\frac{(2m+1)!}{(m!)^2}\,[F(x)\{1 - F(x)\}]^m\,p(x). \tag{12.18}$$

[1)] Выражение (12.12) получается после логарифмирования равенства (12.11) и последующего почленного дифференцирования обеих его частей. — *Прим. перев.*

В общем случае, $100p$%-й квантилью является статистика $X'_{(n+1)p}$, определенная только при целом значении $(n+1)p$. Медиана отвечает $p = \frac{1}{2}$; при $p = \frac{1}{4}$ и $p = \frac{3}{4}$ получаем, соответственно, *нижнюю* и *верхнюю квартили*.

Часто при некоторых условиях регулярности можно получить полезные аппроксимации для моментов порядковых статистик, выраженные через общую плотность распределения величин X_j. Этот подход использует тот факт, что случайные величины $Y_1 = F(X_1)$, $Y_2 = F(X_2), \ldots, Y_n = F(X_n)$ независимы и имеют одинаковое равномерное распределение (см. гл. 26) в промежутке от 0 до 1. Плотность совместного распределения соответствующих порядковых статистик Y'_1, $Y'_2, \ldots, Y'_n$ равна

$$p_{Y'_1,\ldots,Y'_n}(y_1,\ldots,y_n) = n! \qquad (0 \leqslant y_1 \leqslant y_2 \leqslant \ldots \leqslant y_n \leqslant 1)\,.$$

Плотность совместного распределения статистик $Y'_{a_1}, \ldots, Y'_{a_s}$ (для произвольного набора индексов $1 \leqslant a_1 < a_2 < \ldots < a_s \leqslant n$) получается с помощью формулы (12.14) и имеет следующий вид:

$$p_{Y'_{a_1},\ldots,Y'_{a_s}}(y_{a_1},\ldots,y_{a_s}) = \frac{n!}{\prod_{j=1}^{s+1}(a_j - a_{j-1} - 1)!} \prod_{j=1}^{s+1} \left(y_{a_j} - y_{a_{j-1}}\right)^{a_j - a_{j-1} - 1}\,. \tag{12.19}$$

Здесь $a_0 = 0$, $a_{s+1} = n+1$; $y_{a_0} = 0$, $y_{a_{s+1}} = 1$. Простые и смешанные моменты статистик Y'_{a_i} задаются формулой

$$\mathrm{E}\left[\prod_{j=1}^{s} {Y'_{a_j}}^{r_j}\right] = \frac{n!}{\left(n + \sum\limits_{j=1}^{s} r_j\right)!} \prod_{j=1}^{s} \left\{ \frac{\left(a_j + \sum\limits_{i=1}^{j} r_j - 1\right)!}{\left(a_j + \sum\limits_{i=1}^{j-1} r_j - 1\right)!} \right\}. \tag{12.20}$$

Разложим в ряд функцию $X'_r = F^{-1}(Y'_r)$ в окрестности точки $\mathrm{E}\left[Y'_r\right] = r/(n+1)$:

$$X'_r = F^{-1}\left(\frac{r}{n+1}\right) + \left(Y'_r - \frac{r}{n+1}\right)\left[\frac{dF^{-1}(y)}{dy}\bigg|_{y=r/(n+1)}\right] +$$

$$+\frac{1}{2}\left(Y'_r - \frac{r}{n+1}\right)^2 \left[\frac{d^2F^{-1}(y)}{dy^2}\bigg|_{y=r/(n+1)}\right] + \ldots\,. \tag{12.21}$$

Теперь можно взять математическое ожидание от обеих частей равенства (12.21) (используя метод статистических дифференциалов, описанный в гл. 1). Для этого заметим, что так как

$$y = F(x) = \int\limits_{-\infty}^{x} p(t)\, dt,$$

[. . .]