Дорогие студенты, вы зашли в так называемую "Виртуальную аудиторию," в которой к каждому понедельнику я буду заносить материал новых лекций. Если у вас возникают какие-либо вопросы, то вы можете присылать их мне по электронной почте

volodinstudent@gmail.com

На ваши вопросы я буду отвечать вам reply'ем. В конце текста каждой лекции будут предложены задания, которые вам следует выполнить и выслать мне по почте (или, по крайней мере, сообщать, что вы "на проводе"). О том, как их выполнить я могу вам рассказать при общении в Teams'е.

С надеждой на скорое закрытие карантина ваш преподаватель Игорь Николаевич.

Лекция 9

Вероятностные модели роста

Вероятностные модели роста. Условимся употреблять терминологию, связанную с биологическими исследованиями; о приложениях к другим областям естествознания поговорим ниже, после вывода основного уравнения модели.

Предположим, что мы посадили с вами маленькое деревце (саженец) высоты x_0, и во все последующие годы производим замеры $x_1,\, x_2, \ldots$ высоты растущего дерева. Нас интересует прогноз высоты дерева по истечении n лет. Естественно, на ежегодный прирост высоты действует огромное количество природных факторов: температура, осадки, солнечное освещение, плодородие почвы и т.п., поэтому мы, очевидно, имеем дело со стохастическим прогнозом, который формулируется, примерно, как следующее заключение: "Через 60 лет с вероятностью 0,9 высота дерева будет не меньше 15 метров." Конечно, такой прогноз, как и в случае однократного подбрасывания монеты, нельзя применить к одному посаженному дереву, но его можно использовать в прогнозе "зрелости" лесной посадки, состоящей из большого числа деревьев, и тогда наше заключение будет относиться приблизительно к 90% саженцев.

Итак, мы должны трактовать замеры $x_1,\, x_2, \ldots$ в терминах реализаций компонент последовательности случайных величин $\xi_1,\, \xi_2, \ldots$ и попытаться формализовать в математических терминах причину "разброса" в значениях ежегодных приращений $\Delta_k = \xi_k - \xi_{k-1}$ высоты дерева. Естественно предположить, что прирост Δ_k вызван суммарным действием всех тех причин роста, о которых мы говорили выше, то есть действием некоторого неотрицательного "импульса" η_k ($\geqslant 0$). Между Δ_k и η_k существует приближенная линейная связь $\Delta_k = \alpha_k \eta_k$, где α_k зависит от высоты ξ_{k-1} дерева, которой оно достигло по истечении k лет. Положим $\alpha_k = g(\xi_{k-1})$ с естественным условием неотрицательности и непрерывности функции $g(\cdot)$. Таким образом, мы приходим к рекурентным соотношениям, которые описывают ежегодный прирост высоты дерева,

$$\xi_k - \xi_{k-1} = \eta_k g(\xi_{k-1}), \quad k = 1,\, 2, \ldots \tag{1}$$

Нам осталось только сделать некоторые предположения, касающиеся распределения случайных величин $\eta_k,\ k = 1,\, 2, \ldots$. Будем считать, что эти случайные величины неотрицательны, независимы, одинаково распределены и обладают конечными моментами второго порядка: средним значением $a = \mathbf{E}\, \eta_k$ и дисперсией $b^2 = \mathbf{D}\, \eta_k$.

Напомним, что мы интересуемся распределением случайной величины ξ_n, реализация x_n которой указывает размер конкретного дерева по истечении n лет. Перепишем первые n рекурентных соотношений (1) в виде

$$\eta_k = \frac{\xi_k - \xi_{k-1}}{g(\xi_{k-1})}, \quad k = 1, \ldots, n$$

и просуммируем левые и правые части этих равенств. В результате получим

$$\sum_1^n \eta_k = \sum_1^n \frac{\xi_k - \xi_{k-1}}{g(\xi_{k-1})}.$$

Если каждый импульс вызывает незначительный прирост дерева, то есть все $\Delta_k = \xi_k - \xi_{k-1}$ малы, то, трактуя правую часть последнего равенства как интегральную сумму, получаем приближенное равенство

$$\sum_1^n \eta_k = \int_{x_0}^{\xi} \frac{dt}{g(t)}, \tag{2}$$

где $\xi = \xi_n$ – окончательный размер дерева.

Так как функция $g(x)$ положительна, то интеграл в правой части (2) представляет собой некоторую монотонно возрастающую функцию $h(\xi)$. Применение центральной предельной теоремы 14.2 к левой части (2) приводит к утверждению: по истечении достаточно большого срока после посадки дерева ($n \gg 1$) распределение его высоты ξ определяется соотношением $h(\xi) \sim \mathcal{N}(\mu, \sigma^2)$, где $\mu = na$, $\sigma^2 = nb^2$. В силу монотонности функции $h(\cdot)$

$$F(x) = P(\xi < x) = P(h(\xi) < h(x)) = \Phi\left(\frac{h(x) - \mu}{\sigma}\right).$$

Осталось решить проблему с выбором функции $g(\,\cdot\,)$. Если постулировать, что прирост высоты дерева пропорционален достигнутой высоте. то есть положить $g(t) = t$, а именно такое предположение наиболее часто используется в моделях роста, то мы придем к следующему распределению случайной величины ξ.

Логарифмически-нормальное распределение $\mathcal{LN}(\mu, \sigma^2)$.
При $g(t) = t$ интеграл в правой части (2) с точностью до постоянного слагаемого $-\ln x_0$ равен $\ln \xi$, так что $\ln \xi \sim \mathcal{N}(\mu, \sigma^2)$, и функция распределения ξ

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right), \quad x > 0;$$

функция плотности

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}\, x} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}.$$

Это унимодальное, резко асимметричное ($\gamma_1 > 0$) распределение, имеющее очень тяжелый хвост.

Конечно, в практических приложениях целесообразнее оперировать не с ξ, а с его натуральным логарифмом $\eta = \ln \xi$, после чего производить все расчеты, используя модель нормального распределения.

Логарифмически-нормальный закон носит достаточно универсальный характер. Этому распределению подчиняется размер трещины в испытуемом образце материала, который подвергается циклическим нагружениям "на изгиб" – вы можете сами без особых фантазийных усилий пересказать наши построения с высотой дерева в терминах размера трещины. Аналогичные рассуждения могут быть также применены в изучении роста доходов у отдельных лиц достаточно однородной человеческой популяции. Проводимые в этом направлении статистические исследования указывают на хорошее согласие с логарифмически-нормальным распределением достаточно низких доходов, в то время как для умеренных и высоких доходов более подходящим является распределение Парето.

Распределение Парето $\mathrm{Par}(a,\,\alpha)$. Налоговые органы обычно интересуются распределением годовых доходов тех лиц, годовой доход которых превосходит некоторый предел a, установленный законами о налогообложении. Такого рода распределения иногда считают (к сожалению, без особого "экономического" обоснования) приближенно совпадающими с *распределением Парето*, вся вероятностная масса которого сосредоточена в области $x > a$ (носитель распределения $\mathcal{X} = [\,a,\,\infty\,]$), и функция распределения на сегменте $\mathcal{X}$ равна

$$F(x) = 1 - \left(\frac{a}{x}\right)^{\alpha}, \quad x > a,\ \alpha > 0.$$

Это распределение, зависящее от двумерного параметра $\theta = (a,\,\alpha)$ с параметрическим пространством $\Theta = \mathrm{R}_+ \times \mathrm{R}_+$, принадлежит непрерывному типу; его функция плотности в области $x > a$ равна

$$f(x\,|\,\theta) = \frac{\alpha}{a}\left(\frac{a}{x}\right)^{\alpha+1}.$$

Момент k-го порядка у распределения Парето существует только при значениях параметра $\alpha > k$, например, неравенство $\alpha > 1$ гарантирует существование среднего значения, которое, как нетрудно подсчитать, равно $\alpha a/(\alpha - 1)$.

Если случайная величина ξ распределена по закону Парето, то, как легко видеть, $\ln \xi$ имеет показательное распределение, "сдвинутое вправо" на величину $\ln a$, так как

$$P(\ln \xi < x) = P(\xi < e^x) = F(e^x) = 1 - \exp\{\,-\alpha(x - \ln a)\,\}.$$

Это замечание объясняет, почему распределение Парето адекватно описывает распределение наблюдаемых доходов у лиц с высоким уровнем дохода. Вспомним постулат "отсутствия последействия", приводящий к показательному распределению долговечности: вероятность того, что изделие прослужит промежуток времени, не меньший s, при условии, что оно уже отработало срок t, не зависит от величины t. В основу модели Парето положен тот же принцип, только в мультипликативной, а не в аддитивной, формулировке: *вероятность того, что доход отдельного лица увеличится не меньше, чем в s раз, при условии, что он уже достиг уровня t, не зависит от величины достигнутого уровня.* Это происходит, по–видимому, от того, что обладающий большими доходами стремится сохранить достигнутое положение и редко стремится вкладывать большие капиталы в новые отрасли с целью наращивания денежной массы. В таком случае изменчивость дохода за наблюдаемые периоды времени носит случайный характер и не связана с величиной капитала, которым располагают отдельные субъекты. В то же время у "предпринимателей" распределение доходов отлично от закона Парето. Это так называемое логарифмически нормальное распределение.

В рамках построенной нами модели роста часто возникает задача, которую можно трактовать как некоторую альтернативу к проблеме вывода распределения размера, достигнутого к определенному сроку "растущим" объектом исследования. Пусть фиксирован некоторый уровень x размера (дерева, трещины, дохода) и нас интересует распределение момента времени (номера цикла), на котором этот размер будет достигнут. Удивительно, что в рамках нашей модели это распределение не зависит от выбора положительной функции $g(\cdot)$, и получить его можно путем следующих тривиальных рассуждений.

Распределение Бирнбаума–Саундерса BS(λ, θ). Пусть τ – случайная величина, реализующая момент достижения заданного размера x. Тогда событие $\tau > n$ эквивалентно событию $\xi_n < x$ (напомним, все $\eta_k \geqslant 0$) – к моменту времени n высота дерева еще не достигла уровня x. Итак,

$$P(\tau > n) = P(\xi_n < x) = P(h(\xi_n) < h(x)) = \Phi\left(\frac{h(x) - na}{b\sqrt{n}}\right). \tag{3}$$

Заменим теперь n на "непрерывную" переменную t и введем новые параметры λ и θ, определив их уравнениями $\lambda\sqrt{\theta} = h(x)/b, \;\; \lambda/\sqrt{\theta} = a/b$. Цепочка равенств (3) позволяет нам записать распределение случайного момента времени τ, в который дерево (трещина, доход,) достигнет заданного уровня x :

$$F(t) = P(\tau < t) = 1 - \Phi\left(\lambda\left(\sqrt{\frac{\theta}{t}} - \sqrt{\frac{t}{\theta}}\right)\right), \quad t > 0.$$

Это унимодальное распределение, которое называется *распределением Бирнбаума–Саундерса*, и мы будем обозначать его BS(λ, θ). График плотности BS-распределения очень похож на функцию плотности гамма-распределения. BS-распределение играет большую роль при расчетах надежности объектов, долговечность которых определяется развитием трещин, приводящих к гибели объекта.

Разрушение авиационных дисков

12 крепежных отверстий – n=12 испытаний Бернулли.

α – вероятность плохого изготовления, $1-\alpha$ – хорошего,
β – вероятность плохой эксплуатации, $1-\beta$ – хорошей.
Для каждого диска 4 состояния: (ij), $i = 0,\, 1$ (плохое,хорошее изготовление); $j = 0,\, 1$ (плохая,хорошая эксплуатация).
Вероятность состояния (ij) : $Q_{ij} = \alpha^i(1-\alpha)^{1-i}\beta^j(1-\beta)^{1-j}$, $(ij) = (00),\ (01),\ (10),\ (11)$.

$p_{ij} = p_{ij}(t)$ – вероятность наблюдения трещины у любого крепежного отверстия в момент времени t для диска с состоянием (ij). Это функция распределения случайной величины τ – момента появления трещины: $p_{ij}(t) = \mathbf{P}(\tau < t)$.

Пусть ν – случайная величина, значение $k\,(= 0, 1, \ldots, 12)$ которой в момент наблюдения равно числу крепежных отверстий, имеющих трещину. Ее распределение

$$\mathbf{P}\{\nu = k \,|\, \tau < t) = \sum_{i=0}^{1}\sum_{j=0}^{1} Q_{ij}\, \mathrm{C}_{12}^{k}\, p_{ij}^{k}(1-p_{ij})^{12-k}, \quad k = 0, 1, \ldots, 12.$$

Вероятностная модель наблюдения трещины крепежного отверстия в момент t :

$$p_{ij}(t) = \mathbf{P}(\tau < t) = 1 - \Phi\left[\lambda_{ij}\left(\sqrt{\theta_{ij}/t} - \sqrt{t/\theta_{ij}}\right)\right].$$