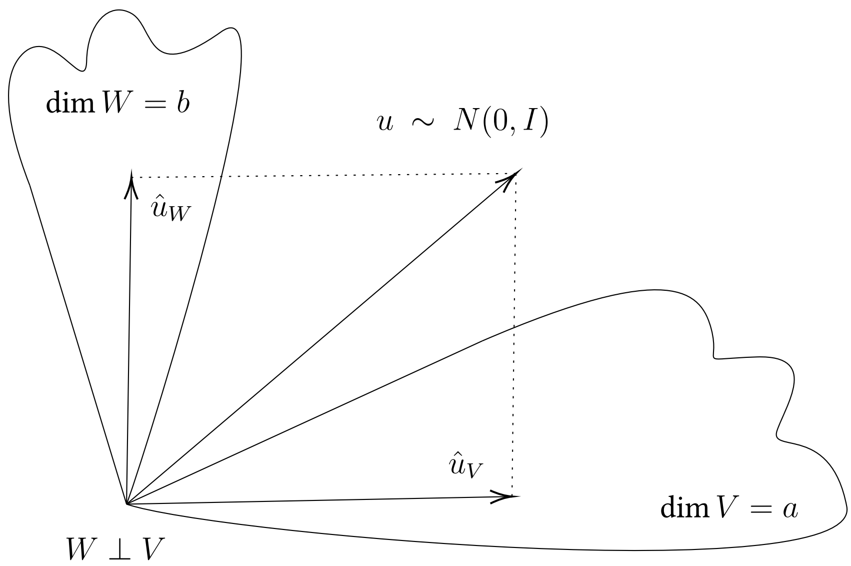
Конспектировали: Любовь Корж и Кирилл Пак

1. Мирок проекций

Случайный вектор u имеет многомерное нормальное стандартное распределение



Такой вектор удовлетворяет ряду аксиом.
Аксиомы:

1. Закон распределения не зависит от угла (а только от длины)
2. u_i независимы
3. $\operatorname{Var}(u_i) = 1$

Из этих аксиом следует два важных текстовых свойства нормального распределения.
Свойства:

1. $\|\hat{u}_V\|^2 \sim \chi^2_a$
2. Если $V \perp W$, то $\|\hat{u}_V\|^2$ и $\|\hat{u}_W\|^2$ независимы

$$\frac{\|\hat{u}_V\|^2/a}{\|\hat{u}_W\|^2/b} \sim \mathcal{F}_{a,b}$$

2. Мирок потока происшествий

Предпосылки:

1. Количество происшествий за разные интервалы времени независимы
2. Вероятность того, что за малый интервал времени произойдёт ровно одно (хотя бы одно) событие, пропорциональна длине интервала

$$P(\text{событие в интервале } [t; t+\Delta]) = \lambda\Delta + o(\Delta)$$

Следствия:

1. Время между соседними событиями имеет экспоненциальное распределение $x_i \sim exp(\lambda)$

$$f(x) = \lambda \exp(-\lambda x)$$

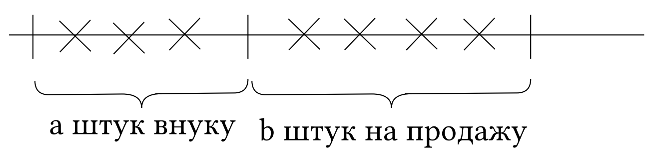
2. Время на n происшествий $x_1 + x_2 + ... + x_n = S_n \sim Gamma(n, \lambda)$

$$f(s) = const \cdot s^{n-1} e^{-\lambda s}$$

$$\mathrm{E}(S_n) = n \cdot \frac{1}{\lambda}$$

$$\mathrm{Var}(S_n) = n \cdot \mathrm{Var}(X_1) = n \cdot \frac{1}{\lambda^2}$$

3. Вспомним задачку про тётю Мотю, пекущую блинчики внуку и на продажу



$$A = \frac{S_a}{S_{a+b}} \text{ — доля времени на блинчики внуку от общего времени}$$

$$A \in [0; 1]$$

$$\mathrm{E}[A] = \frac{a}{a+b}$$

$$\frac{S_a}{S_{a+b}} \sim Beta(a, b) \text{ (по определению)}$$

Частный случай:
$a = 2$ (внуку)
$b = 1$ (на продажу)
Время на три блинчика соответственно равно x_1, x_2, x_3

$$A_1 = \frac{X_1}{X_1 + X_2} \sim Beta(1, 1)$$

$$A_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3} \sim Beta(2,1)$$
$$S_3 = X_1 + X_2 + X_3 \sim Gamma(3, \lambda)$$

Выразим x_1, x_2, x_3 через a_1, a_2, s_3:

$$x_1 = a_1 \cdot a_2 \cdot s_3$$
$$x_2 = (1 - a_1) \cdot a_2 \cdot s_3$$
$$x_3 = (1 - a_1) \cdot s_3 = s_3 - a_2 \cdot s_3$$

Посчитаем $f(x_1, x_2, x_3) \cdot dx_1 \wedge dx_2 \wedge dx_3$
Отдельно вычислим $dx_1 \wedge dx_2 \wedge dx_3$:

$$\begin{aligned} dx_1 \wedge dx_2 = d(a_1 a_2 s_3) \wedge d((1 - a_1) a_2 s_3) = \\ = d(a_1 a_2 s_3 \wedge (d(a_2 s_3) - d(a_1 a_2 s_3)) = \\ = d(a_1 a_2 s_3) \wedge d(a_2 s_3) = (da_1 a_2 s_3 + a_1 d(a_2 s_3) \wedge d(a_2 s_3) = \\ a_2 s_3 da_1 \wedge d(a_2 s_3) \end{aligned}$$

$$\begin{aligned} dx_1 \wedge dx_2 \wedge dx_3 = a_2 s_3 da_1 \wedge d(a_2 s_3) \wedge d(s_3 - a_2 s_3) = \\ = a_2 s_3 da_1 \wedge (da_2 s_3 + a_2 ds_3) \wedge ds_3 = \\ = a_2 s_3^2 \cdot da_1 \wedge da_2 \wedge ds_3 \end{aligned}$$

$$\underbrace{\lambda^3 \cdot e^{-\lambda s_3} \cdot a_2 s_3^2}_{\text{совместная ф-ия плотности}} \quad da_1 \wedge da_2 \wedge ds_3$$

$$\underbrace{?}_{f(a_1)} \cdot \underbrace{? \cdot a_2}_{f(a_2)} \cdot \underbrace{? \cdot s_3^2 e^{-\lambda s_3}}_{f(s_3)}$$

$$1 \cdot 2a_2 \cdot \frac{\lambda^3}{2} s_3^2 e^{-\lambda s_3}$$

$$dx_1 \wedge dx_2 \wedge dx_3 = a_2 s_3^2 da_1 \wedge da_2 \wedge ds_3$$

Если печь внуку 3, в не 2 блинчика, то добавится $A3 = \frac{X_1+X_2+X_3}{X_1+X_2+X_3+X_4}$, s_3 превратится в $a_3 s_4$ и x_4 в $s_4 - s_3$

$$\begin{aligned} dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4 = a_2 s_3^2 da_1 \wedge da_2 \wedge ds_3 \wedge d(s_4 - s_3) = \\ = a_2 s_3^2 da_1 \wedge da_2 \wedge ds_3 \wedge ds_4 = \\ = a_2 (a_3 s_4)^2 da_1 \wedge da_2 \wedge d(a_3 s_4) \wedge ds_4 = \\ = a_2 a_3^2 s_4^2 da_1 \wedge da_2 \wedge (da_3 s_4 + ds_4 a_3) \wedge ds_4 = \\ = a_2 a_3^2 s_4^2 da_1 \wedge da_2 \wedge s_4 da_3 \wedge ds_4 = \\ = a_2 \cdot a_3^2 \cdot s_4^3 \wedge da_1 \wedge da_2 \wedge da_3 \wedge ds_4 \end{aligned}$$

Получается, что при изготовлении a блинчиков внуку и 1 на продажу

$$f(x) = ? x^{a-1}$$

Аналогично, при изготовлении 1 блинчика внуку и b на продажу

$$f(x) = ?(1-x)^{b-1}$$

Логично предположить, что при изготовлении a блинчиков внуку и b блинчиков на подажу

$$f(x) = ?x^{a-1}(1-x)^{b-1}$$

Мораль:
Если X — доля времени, которое тратится на испечение a блинчиков, если всего пеку $a+b$ блинчиков, то $x \in [0;1)$, $X \sim Beta(a,b)$ и $f_{a,b}(x) = ?x^{a-1}(1-x)^{b-1}$

3. Вывод функции плотности χ^2-распределения. Связь χ^2 и гамма распределений

Рассмотрим стандартную нормальную случайную величину $X \sim \mathcal{N}(0,1)$. Мы знаем, как выглядит ее функция распределения:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Давайте выведем функцию плотности случайной величины, которая имеет χ^2-распределение с 1 степенью свободы. По определению случайная величина $Q = X^2 \sim \chi^2_1$ и представляет собой квадрат длины вектора из одной компоненты. Таким образом, нам надо найти функцию плотности распределения случайной величины Q.

Воспользуемся дифференциальной формой X (техника с "птичками"):

$$f(x)dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Сделаем замену. Так как $X = \pm\sqrt{Q}$, то $x = \pm\sqrt{q}$. Заметим, что для получения дифференциальной формы для Q надо подставить в дифференциальную форму для X две точки $x = \sqrt{q}$ и $x = -\sqrt{q}$, что в силу симметричности функции плотности распредения X означает просто домножение на 2:

$$2f(\sqrt{q})d\sqrt{q} = 2\frac{1}{\sqrt{2\pi}} e^{-\frac{q}{2}} \frac{1}{2\sqrt{q}} dq = \frac{1}{\sqrt{2\pi}} q^{-\frac{1}{2}} e^{-\frac{q}{2}} dq$$

То, что стоит перед дифференциалом, и есть функция плотности случайной величины $Q \sim \chi^2_1$:

$$f(q) = \frac{1}{\sqrt{2\pi}} q^{-\frac{1}{2}} e^{-\frac{q}{2}}$$

Вспомним, что функция плотности гамма-распределения имеет вид:

$$f(s) = const \times s^{n-1} e^{-\lambda s}$$

Таким образом, мы с чистой совестью заключаем, что χ^2-распределение с 1 степенью свободы, — это ни что иное, как гамма-распредение с параметрами $n = \frac{1}{2}$ и $\lambda = \frac{1}{2}$:

$$\chi^2_1 \sim Gamma\left(n = \frac{1}{2}, \lambda = \frac{1}{2}\right)$$

Получается, что χ^2-распределение — это частный случай гамма-распределения. Например, случайная величина, имеющая χ^2-распределение с 2 степенями свободы, представляет собой сумму двух случайных величин, имеющих χ^2-распределение с 1 степенью свободы, и значит имеет гамма-распредение с параметрами $n = 1$ и $\lambda = \frac{1}{2}$:

$$\chi^2_2 \sim \chi^2_1 + \chi^2_1 \sim Gamma\left(n = \frac{1}{2}, \lambda = \frac{1}{2}\right) + Gamma\left(n = \frac{1}{2}, \lambda = \frac{1}{2}\right) \sim$$
$$\sim Gamma\left(n = 1, \lambda = \frac{1}{2}\right) \sim Exp\left(\frac{1}{2}\right)$$

Т.к. гамма-распределение с $n = 1$ — это экспоненциальное распределение, то в итоге χ^2-распределение с 2 степенями свободы можно трактовать (в терминах тети Моти, которая печет блинчики), как время ожидания одного блинчика, если в среднем 1 блинчик печется раз в 2 минуты (или, например, 2 часа, в зависимости от единиц измерения времени).

4. Вспоминая теорему Гаусса-Маркова...

Теперь мы можем сформулировать несколько теорем.

Теорема 1. Если:

- [ТГМ1] Предполагаем, что верна модель $y = X\beta + u$.
- [ТГМ2] Оцениваем модель регрессии $\hat{y} = X\hat{\beta}$ с помощью МНК.
- [ТГМ3] β — константы.
- [ТГМ4] X стохастистические $(n > k)$:

$$P(X \text{ имеет полный ранг}) = 1$$

- [ТГМ5] $\mathrm{E}(u|X) = 0, \mathrm{Var}(u|X) = \sigma^2 I$
- [нормальность] $u|X \sim \mathcal{N}(0, \sigma^2 I)$

то:

- $\dfrac{RSS}{\sigma^2} \sim \chi^2_{n-k} \sim Gamma\left(\dfrac{n-k}{2}, \lambda = \dfrac{1}{2}\right)$
(новизна: гамма-распределение)

Теорема 2. Если:

- [ТГМ1'] Предполагаем, что верна модель $y = \beta_1 \mathbf{1} + \mathbf{u}$.
- [ТГМ2] По-прежнему оцениваем модель регрессии $\hat{y} = X\hat{\beta}$ с помощью МНК.
- Выполнены предпосылки [ТГМ3-5] и [нормальность]

то:

- $\dfrac{RSS}{\sigma^2} \sim \chi^2_{n-k} \sim Gamma\left(\dfrac{n-k}{2}, \lambda = \dfrac{1}{2}\right)$
  (как и было)
  но в дополнение:

- $\dfrac{ESS}{\sigma^2} \sim \chi^2_{k-1} \sim Gamma\left(\dfrac{k-1}{2}, \lambda = \dfrac{1}{2}\right)$

- $\dfrac{ESS/(k-1)}{RSS/(n-k)} \sim \mathcal{F}_{k-1,n-k}$
  (позволяет проверить гипотезу: $H_0 : \beta_2 = \beta_3 = ... = \beta_k = 0$)

- $\dfrac{TSS}{\sigma^2} \sim \chi^2_{n-1} \sim Gamma\left(\dfrac{n-1}{2}, \lambda = \dfrac{1}{2}\right)$

- $R^2 = \dfrac{ESS}{TSS} = \dfrac{ESS}{ESS + RSS} \sim Beta\left(\dfrac{k-1}{2}, \dfrac{n-k}{2}\right)$
  (выполнено в силу того, что $\dfrac{ESS}{\sigma^2}$ и $\dfrac{RSS}{\sigma^2}$ независимые случайные величины, имеющие гамма-распределение)

Знание о том, что R^2 имеет бета-распределение позволяет, например, найти его математическое ожидание:

$$\mathrm{E}(R^2) = \frac{\left(\frac{k-1}{2}\right)}{\frac{k-1}{2} + \frac{n-k}{2}} = \frac{k-1}{n-k}$$

Теорема 3. Если:

- [ТГМ1”] Предполагаем, что верна модель $y = X\beta + u$, но часть коэффициентов занулена:

$$\beta = \begin{pmatrix} \beta_{R+1} \\ \vdots \\ \beta_R \\ \beta_{R+1} \\ \vdots \\ \beta_{UR} \end{pmatrix}$$

  Предполагаем $H_0 : \beta_{R+1} = \beta_{R+2} = ... = \beta_{UR} = 0$

- [ТГМ2”] Оцениваем 2 модели регрессии: короткую (R, в предположении H_0) и длинную (UR, $\hat{y} = X\hat{\beta}$, без ограничений) с помощью МНК. В результате получаем блок показателей по длинной модели ($\hat{y}^{UR}, \hat{\beta}^{UR}, RSS_{UR}, ESS_{UR}$) и по короткой модели ($\hat{y}^R, \hat{\beta}^R, RSS_R, ESS_R$). TSS у них общий, т.к. не зависит от выбранной модели.

- Выполнены предпосылки [ТГМ3-5] и [нормальность]

то:

- $\dfrac{(RSS_R - RSS_{UR})/k_{UR} - k_R)}{RSS_{UR}/(n - k_{UR})} \sim \mathcal{F}_{k_{UR}-k_R, n-k_{UR}}$
  (в силу независимости RSS_{UR} и $RSS_R - RSS_{UR}$)
- $\dfrac{RSS_{UR}}{RSS_R} = \dfrac{RSS_{UR}}{RSS_{UR} + (RSS_R - RSS_{UR})} \sim Beta\left(\dfrac{n - k_{UR}}{2}, \dfrac{k_{UR} - k_R}{2}\right)$
  (Заметим, что RSS_{UR} и RSS_R зависимы, а вот RSS_{UR} и $RSS_R - RSS_{UR}$ независимы)
- $\dfrac{RSS_R - RSS_{UR}}{RSS_R} \sim Beta\left(\dfrac{k_{UR} - k_R}{2}, \dfrac{n - k_{UR}}{2}\right)$
  (Заметим, что $RSS_R - RSS_{UR}$ и RSS_R зависимы)
- $1 - R^2 = \dfrac{RSS}{RSS + ESS} \sim Beta\left(\dfrac{n-k}{2}, \dfrac{k-1}{2}\right)$