1 Обязательное домашнее задание

В этот раз (и, вероятно, так будет и далее) задание состоит из двух частей, "бумажной" и "прогерской". Соответственно, на следующем занятии нужно будет решить одну на мой выбор задачу из предложенных бумажных, а код решения прогерской задачи заранее прислать мне на почту (можно прислать ссылку на github/bitbucket).

Код можно писать на любом разумном языке, но я бы предпочел R, Python или C++. Для R необходимые средства моделирования уже встроенны в язык (справку по ним можно получить, дав команду `?Distributions`), для Python можно использовать стандарный модуль `random` (также может быть полезным `numpy`), для C++ — модуль `std::random` из нового стандарта или `boost::random`.

1.1 Прогерская задача

Сама задача следующая:

Задание 1.1. Рассмотрим интеграл $\int_0^1 \sin(x+e^{-x})dx$. С помощью метода Монте-Карло убедиться что его значение приближенно равно 0.8997.

Никаких доверительных интервалов строить не нужно, достаточно продемонстрировать поведение среднего

Есть усложненный вариант этой задачи (ниже по тексту)

1.2 Бумажные задачи

Задание 1.2. Пусть $\xi_1, \ldots, \xi_d \sim \mathcal{N}(\mu, \sigma^2)$, независимы. Найти распределение с.в. $\eta = \sum_{i=1}^d \xi_i/d$, т.е. распределение выборочного среднего.

Задание 1.3. Пусть $\alpha \sim U[0, 1], n \in \mathsf{N}$. Найти распределения величин $\lceil n\alpha \rceil$ и $\{n\alpha\}$. *Целая и дробная часть* Будут ли они независимы?

Задание 1.4. Пусть $\xi_1, \xi_2 \sim U[0, 1]$ — независимые равномерно распределенные с.в. Найти распределение случайной величины $\eta = \xi_1 + \xi_2$.

Задание 1.5. Пусть распределение случайной величины ξ — смесь d распределений с плотностями $f_1(x), \ldots, f_d(x)$ взятых с весами $p_1, \ldots, p_k$. Пусть в случайном эксперименте получено значение $\xi = a$. К какой из компонент смеси наиболее вероятно (в смысле байесовской вероятности) относится полученное наблюдение? *Сюжетно задача выглядит так. Пусть случайно взятый работник с некоторой вероятностью (p_1) оказывается рабочим цеха (в противном случае он ИТР). Пусть нам известны распределения (f_1, f_2) роста работников в каждой группе. Мы измеряем рост работника (a) и хотим понять, к какой группе он относится (более вероятно относится).*

Задание 1.6. Для стандартного нормального $\mathcal{N}(0, 1)$ распределения найти квантили уровня $0.5, 0.95, 0.975, 0.99$ и 0.995.

Задание 1.7. Для произвольного нормального распределения найти значения асимметрии и экцесса.

2 Задачи, разобранные 21 марта

Задание 2.1. Пусть $\alpha \sim U[0, 1]$. Найти условное распределение $\alpha \mid \alpha > a$.

Задание 2.2. Пусть $\xi \sim Geom(\lambda)$ (геометрическое). Найти условное распределение $\xi - a \mid \xi > a$.

Задание 2.3. Пусть f — некоторая произвольная функция. Найти матожидание и дисперсию (если они существуют) величины $f(\xi)/g(\xi)$, где ξ — случайная величина с плотностью $g(x)$. При каких условиях существуют матожидание и дисперсия?

Задание 2.4. Пусть ξ, η — независимые случайные величины с плотностями $f_\xi(x), f_\eta(x)$ соответственно. Найти плотность с.в. $\xi + \eta$.

Задание 2.5. Пусть $\xi_1, \xi_2, \ldots, \xi_d \sim \Gamma(\theta, k_i)$ — независимые случайные величины с Гамма-распределением. *θ одна на всех, а k могут быть разными* Найти распределение случайной величины $\eta = \sum_i \xi_i$.

Задание 2.6. Пусть $\xi_1, \xi_2, \ldots, \xi_d$ — независимые нормально распределенные случайные величины с матожиданиями μ_i и дисперсиями σ_i^2. Найти распределение $\eta = \sum_i \xi_i$.

Решение. Ответ: $\eta \sim \mathcal{N}(\sum_i \mu_i, \sum_i \sigma_i^2)$.

Не умаляя общности, сделаем три упрощения. (1) Нам достаточно доказать результат для $d = 2$, а далее по индукции. (2) Можно считать, что $\mu_1 = \mu_2 = 0$, так как параметр μ имеет смысл сдвига, т.е. если $\xi \sim \mathcal{N}(\mu, \sigma^2)$, то $\xi + a \sim \mathcal{N}(\mu + a, \sigma^2)$. Это следует, например, из формулы изменения плотности при линейном преобразовании. (3) Из нее же следует, что мы можем считать $\sigma_1^2 = 1$. σ_2^2 (не обязательно равную 1) обозначим за σ^2.

$$f_{\xi_1+\xi_2}(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-t)^2}{2\sigma^2}} dt =$$

Константный множитель обозначим за C, он не будет нас интересовать, т.к. это просто нормировка.

$$= C \int_{-\infty}^{\infty} e^{-\frac{t^2\sigma^2 + (x-t)^2}{2\sigma^2}} dt = C \int_{-\infty}^{\infty} e^{-\frac{t^2(\sigma^2+1) - 2tx + x^2}{2\sigma^2}} dt =$$

$$= C \int_{-\infty}^{\infty} e^{-\frac{t^2(\sigma^2+1) - 2tx + x^2/(\sigma^2+1) - x^2/(\sigma^2+1) + x^2}{2\sigma^2}} dt = C \int_{-\infty}^{\infty} e^{-\frac{\left(t\sqrt{\sigma^2+1} - x/\sqrt{\sigma^2+1}\right)^2 + x^2\left(1 - (\sigma^2+1)^{-1}\right)}{2\sigma^2}} dt =$$

$$= C \cdot e^{-\frac{x^2\left(1-(\sigma^2+1)^{-1}\right)}{2\sigma^2}} \cdot \int_{-\infty}^{\infty} e^{-\frac{\left(t\sqrt{\sigma^2+1} - x/\sqrt{\sigma^2+1}\right)^2}{2\sigma^2}} dt =$$

Оставшийся интеграл от x на самом деле не зависит (это легко показать, сделав замену переменной; пределы интегрирования при этом не изменятся), следовательно, он уходит в константу.

$$= C' e^{-\frac{x^2\left(1-(\sigma^2+1)^{-1}\right)}{2\sigma^2}} = C' e^{-\frac{x^2(\sigma^2+1-1)}{2\sigma^2(\sigma^2+1)}} = C' e^{-\frac{x^2}{2(\sigma^2+1)}}.$$

Итак, получили требуемую плотность (с точностью до константы, но константа естественно получается из нормировки).

Задание 2.7. Пусть случайная величина η имеет распределение, являющееся смесью распределений случайных величин $\xi_1, \ldots, \xi_d$ с вероятностями (весами) $p_1, \ldots, p_d$. Выразить матожидание, функцию распределения, плотность (если все с.в. ξ_i непрерывны), производящую функцию (если они дискретны) случайной величины η.

3 Более сложные задачи

Задание 3.1. С помощью метода Монте-Карло показать, что:

1. $\int_0^\infty \sin x \cdot e^{-x^2} dx \approx 0.4244$
2. $\int_0^1 (\sin x)^{-2/3} dx \approx 3.0500$

Задание 3.2. Исходя из определения $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ (среднего, стандартного отклонения, асимметрии и эксцесса) и их значений для нормального распределения опеределить γ_k.