Теория вероятностей

Пространство с мерой, такое что мера всего пространства равна 1, называется **вероятностным пространством**. В дальнейшем подразумевается, что вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ фиксировано раз и навсегда. В контексте вероятностных мер некоторые математические объекты имеют особые названия, так как этот предмет развивался независимо от теории меры Лебега. Например:

- *Измеримое множество* называется **событием**.
- *Измеримая функция* X называется **случайной величиной**. Если X принимает значения в $\mathbb{R}$, она называется **вещественной случайной величиной** или просто случайной величиной. В общем случае, если $X\colon \Omega \to E$, то X — это E**-значная случайная величина**. Подразумевается, что E — измеримое пространство.
- Образ меры $\mu := \mathbb{P} \circ X^{-1}$ называется **распределением** X.
- Обозначение *почти всюду* или *п.в.* заменяется на **почти наверное** или **п.н.**
- Используется вероятностная запись: $\mathbb{P}\left(\{\omega \in \Omega \ : \ \text{свойство } p(\omega) \text{ выполнено}\}\right)$ обозначается $\mathbb{P}\left(p\right)$. Например, $\mathbb{P}(X \geq 5)$ означает $\mathbb{P}\left(\{\omega \in \Omega \ : \ X(\omega) \geq 5\}\right)$. Это связано с тем, что вероятностное пространство не является каноническим, и обычно интересуются свойствами, которые выполняются независимо от конкретного выбора вероятностного пространства. Теоретико-мерная настройка используется в основном для обеспечения существования таких пространств.
- Интеграл называется **математическим ожиданием** и обозначается $\mathbb{E}[X] := \int X(\omega) d\mathbb{P}(\omega)$. Например, для E-значной случайной величины X и $f\colon E \to \mathbb{R}$, таких что $f(X) \in L^1(\mathbb{P})$,

$$\mathbb{E}[f(X)] = \int_E f d\mu$$

  где μ — распределение X.

Дискретные вероятности

Пусть $(\Omega, \mathcal{F})$ — измеримое пространство. Вероятность μ на Ω называется *дискретной*, если существует счётное подмножество $S \subset \Omega$, такое что $\mu(S) = 1$. Для дискретной меры интуитивно понятно, что для характеристики вероятности μ достаточно знать вероятность каждой точки $x \in S$.

Действительно, легко устанавливается следующий результат.

Утверждение 0.1. *Существует взаимно однозначное соответствие между*

a. *Множеством дискретных вероятностей $\mathcal{P}_{\text{discrete}}(\Omega)$ на Ω.*
b. *Отображениями $\Omega \ni \omega \mapsto \mu_\omega \in [0,1]$, такими что $\sum_{\omega \in \Omega} \mu_\omega = 1$ (сходимость суммы подразумевает, что множество $\{\omega \ : \ \mu_\omega > 0\}$ не более чем счётно).*

Такое соответствие задается следующим образом:

- *Для $\mu \in \mathcal{P}_{\text{discrete}}(\Omega)$ положим $\mu_\omega := \mu(\{\omega\})$. Тогда $\omega \mapsto \mu_\omega$ удовлетворяет гипотезе из пункта b.*

- *Для отображения $\mu_\cdot$, как в пункте b., определим вероятность на Ω, положив $\mu(A) := \sum_{\omega \in A} \mu_\omega$.*

Ясно, что для счётных пространств любая вероятность дискретна. В этом случае набор $(\mu_\omega)_\omega$ обычно называют вероятностью.

Этот подход легко распространяется на атомические вероятности.

Определение 0.1 (атомические вероятности)**.** Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство. Измеримое множество $A \in \mathcal{F}$ называется **атомом** для $\mathbb{P}$, если $\mathbb{P}(A) > 0$ и не существует измеримого $B \subset A$ с $0 < \mathbb{P}(B) < \mathbb{P}(A)$.

$\mathbb{P}$ называется **атомической**, если каждое множество строго положительной меры содержит атом.

Условное математическое ожидание

Для дискретных случайных величин смысл условных ожиданий и вероятностей вполне интуитивен. Если A — событие с $\mathbb{P}(A) \in (0,1)$, а X — вещественная случайная величина, то

$$\mathbb{E}[X|A] = \mathbb{E}[X\mathbf{1}_A]/\mathbb{P}(A), \qquad \mathbb{E}[X|A^c] = \mathbb{E}[X\mathbf{1}_{A^c}]/\mathbb{P}(A^c) \tag{1}$$

является интуитивным определением.

Однако существует важное понятие, расширяющее этот элементарный подход. В Уравнение 1 мы можем отождествить два значения, даваемые формулой, с функцией на Ω, которая принимает одно постоянное значение на A и другое постоянное значение на A^c. В теоретико-мерных терминах это функция, которая измерима относительно σ-алгебры $\mathcal{G}_A := \{\emptyset, A, A^c, \Omega\}$. Как мы увидим, удобно обозначать эту функцию следующим образом:

$$\mathbb{E}[X|\mathcal{G}_A](\omega) = \begin{cases} \mathbb{E}[X\mathbf{1}_A]/\mathbb{P}(A) & \text{если } \omega \in A \\ \mathbb{E}[X\mathbf{1}_{A^c}]/\mathbb{P}(A^c) & \text{если } \omega \in A^c \end{cases}$$

Тогда случайная величина $Z := \mathbb{E}[X|\mathcal{G}_A]$ обладает двумя свойствами:

a. Как было замечено, Z является $\mathcal{G}_A$-измеримой (как функция из Ω в $\mathbb{R}$).
b. Для любой случайной величины Y, которая также $\mathcal{G}_A$-измерима, скажем $Y(\omega) = \alpha$ на A и $Y(\omega) = \beta$ на A^c, имеет место равенство
$$\mathbb{E}[ZY] = \mathbb{E}[XY] \tag{2}$$

Принимая $\alpha = 1$ и $\beta = 0$ и наоборот, нетрудно проверить, что Z — единственная случайная величина (с точностью до эквивалентности п.н.), обладающая такими свойствами.

Упражнение 0.1. Проверьте, что Уравнение 2 выполняется и что Z определена однозначно с точностью до эквивалентности п.н.

Вышеупомянутые свойства a. и b. — это то, что позволяет нам определить условное ожидание для любой под-σ-алгебры.

Примечание. Пусть $\mathcal{G}$ — под-σ-алгебра $\mathcal{F}$, и пусть $X \in L^1(\mathbb{P})$ — случайная величина. Тогда существует единственная (с точностью до эквивалентности п.н.) случайная величина Z, такая что:

a. Z является $\mathcal{G}$-измеримой.
b. Для всех ограниченных $\mathcal{G}$-измеримых случайных величин Y имеет место равенство

$$\mathbb{E}[ZY] = \mathbb{E}[XY] \tag{3}$$

Действительно, рассмотрим (знакопеременную) конечную меру ν на $(\Omega, \mathcal{G})$, заданную как $\nu(A) := \mathbb{E}[X\mathbf{1}_A]$. Обозначим $\mathbb{P}_\mathcal{G}$ ограничение $\mathbb{P}$ на $\mathcal{G}$.

ν абсолютно непрерывна относительно $\mathbb{P}_\mathcal{G}$ (так как $\nu(A) = 0$, если $\mathbb{P}(A) = 0$). По теореме Радона — Никодима (применённой к положительной и отрицательной частям X), существует единственная функция $Z\colon \Omega \to \mathbb{R}$, которая $\mathcal{G}$-измерима, такая что выполняется условие b.

Определение 0.2. Единственная (с точностью до эквивалентности п.н.) случайная величина Z, определённая в приведенном выше замечании, называется **условным математическим ожиданием X относительно $\mathcal{G}$** (или при условии $\mathcal{G}$).

Утверждение 0.2 (Свойства условного ожидания)**.** *Имеем:*

1. *Если $\mathcal{H}$ — под-σ-алгебра $\mathcal{G}$, то п.н.*
$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$$
В частности, если $\mathcal{H}$ — тривиальная σ-алгебра, $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.
2. *Если X независима от $\mathcal{G}$, то $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.*
3. *Если Y — $\mathcal{G}$-измерима и $XY \in L^1(\Omega)$, то $\mathbb{E}[XY|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]Y$ п.н.*
4. *Теоремы о сходимости интегралов распространяются на их условные версии. Например, если $X_n \uparrow X$ п.н. (монотонная сходимость), то $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$.*

Упражнение 0.2. Докажите Утверждение 0.2.

Упражнение 0.3. Предположим, что $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, и пусть $\mathcal{G} \subset \mathcal{F}$ — под-σ-алгебра $\mathcal{F}$. Докажите, что $\mathbb{E}[X|\mathcal{G}]$ является ортогональной проекцией (в гильбертовом пространстве $L^2(\Omega, \mathcal{F}, \mathbb{P})$) величины X на (замкнутое) подпространство $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Используйте этот факт, чтобы дать непосредственную интерпретацию свойства 1 в Утверждение 0.2.

Топологии на пространстве вероятностных мер

Пусть $(E, \mathcal{E})$ — измеримое пространство. Рассмотрим некоторые распространенные расстояния на пространстве $\mathcal{P}(E)$ вероятностных мер на E.

Определение 0.3. Для $\mu, \nu \in \mathcal{P}(E)$ определим расстояние по **полной вариации** как
$$\|\mu - \nu\|_{TV} := \sup_A |\mu(A) - \nu(A)| = \frac{1}{2} \sup_{|f|\leq 1} \int f\,d\mu - \int f\,d\nu$$
где супремумы берутся по измеримым событиям $A \subset E$ и измеримым функциям $f\colon E \to \mathbb{R}$ с $|f| \leq 1$.

Примечание. Выполняется $\|\mu - \nu\|_{TV} \leq 1$. Более того, если $E = \mathbb{R}$, $\mu = \varrho dx$ и $\nu = \varrho' dx$, то $\|\mu - \nu\|_{TV} = \frac{1}{2}\|\varrho - \varrho'\|_{L^1}$.

Определение 0.4. Предположим, что E — метрическое пространство с расстоянием d, и $\mathcal{E}$ — ассоциированная борелевская σ-алгебра. Для $A \subset E$ и $\varepsilon > 0$ определим $A^\varepsilon := \{x \in E \,:\, d(x, A) < \varepsilon\}$. Для $\lambda > 0$ **метрика Леви — Прохорова** $d_\lambda\colon \mathcal{P}(E) \times \mathcal{P}(E) \to [0, \lambda]$ определяется как
$$d_\lambda(\mu, \nu) := \inf\Big\{\varepsilon > 0 \,:\, \mu(A) \leq \nu(A^\varepsilon) + \frac{\varepsilon}{\lambda}, \nu(A) \leq \mu(A^\varepsilon) + \frac{\varepsilon}{\lambda}, \qquad \text{для всех замкнутых } A \subset E\Big\}$$

Примечание. Имеем:

1. $d_\lambda(\delta_x, \delta_y) = \min(d(x,y), \lambda)$. Другими словами, если λ больше диаметра E, то d_λ действительно является поднятием метрики d на $\mathcal{P}(E)$.
2. Если d и d' порождают одну и ту же топологию на E, то d_λ и $d'_{\lambda'}$ порождают одну и ту же топологию на $\mathcal{P}(E)$.
3. Если E — польское пространство, то $\mathcal{P}(E)$, снабженное метрикой Леви — Прохорова, является польским пространством.
4. Если E — польское пространство, последовательность μ_n сходится к μ в $\mathcal{P}(E)$ (в метрике Леви — Прохорова) тогда и только тогда, когда
$$\lim_n \int f d\mu_n = \int f d\mu$$
для всех $f \in C_\mathrm{b}(E)$.
5. Подмножество $\mathcal{K} \subset \mathcal{P}(E)$ предкомпактно в топологии Леви — Прохорова тогда и только тогда, когда[1]
$$\inf_{K \text{compact}} \sup_{\mu \in \mathcal{K}} \mu(K^c) = 0 \tag{4}$$
6. Мы можем перефразировать последний пункт следующим образом: для последовательности (μ_n) в $\mathcal{P}(E)$ существует подпоследовательность n_k и неотрицательная борелевская мера μ с $\mu(E) \leq 1$, такие что $\mu_{n_k}(f) \to \mu(f)$ для любой $f \in C_\mathrm{b}(E)$. μ является вероятностью тогда и только тогда, когда для каждого $\varepsilon > 0$ существует компакт $K^\varepsilon \subset E$, такой что
$$\varliminf_k \mu_{n_k}(K^\varepsilon) \geq 1 - \varepsilon$$

Энтропии

Понятие энтропии может быть введено в нескольких различных контекстах и с несколько разным смыслом.

Определение 0.5 (Относительная энтропия). Пусть $(E, \mathcal{E}, m)$ — вероятностное пространство с опорной мерой m. **Относительная энтропия** (в математике) или **дивергенция Кульбака — Лейблера** (в информатике) между μ и m равна
$$H(\mu|m) := \sup_f \int f d\mu - \log \int e^f dm$$
где супремум берется по всем ограниченным измеримым функциям $f\colon E \to \mathbb{R}$.

Утверждение 0.3. *Имеем:*

1. $H(\mu|m) \geq 0$ *и* $H(\mu|m) = 0$ *тогда и только тогда, когда* $\mu = m$.
2. $H(\cdot|\cdot)$ *является совместно выпуклой, а именно* $H(\alpha\mu + (1-\alpha)\mu'|\alpha m + (1-\alpha)m') \leq \alpha H(\mu|m) + (1-\alpha)H(\mu'|m')$.
3. *Если* E *— польское пространство,* $H(\mu|m)$ *является полунепрерывной снизу по совокупности аргументов в метрике Леви — Прохорова.*
4. *Для* $h(v) := v \log v$ *(или, эквивалентно,* $h(v) = v\log v - v + 1$*)*

$$H(\mu|m) = \begin{cases} \int h(\varrho) dm & \text{если } \mu = \varrho m \text{ (в смысле Радона — Никодима)} \\ +\infty & \text{в противном случае} \end{cases} \tag{5}$$

[1] Свойство Уравнение 4 обычно называют **плотностью** семейства вероятностей $\mathcal{K}$.

5. *Если $m, m' \in \mathcal{P}(E)$, причем $m = \frac{1}{Z}e^{-V}m'$ для некоторого измеримого $V \in L^1(\mu)$ и $Z > 0$, то*

$$H(\mu|m) = H(\mu|m') + \int V d\mu + \log Z \tag{6}$$

6. *Для каждого события $A \subset E$*

$$\mu(A) \leq \frac{H(\mu|m) + \log 2}{1 + \log(1/m(A))}$$

7. *Выполняется*

$$\begin{aligned} \|\mu - m\|_{TV}^2 &\leq \tfrac{1}{2}H(\mu|m) \\ \|\mu - m\|_{TV}^2 &\leq 1 - \exp(-H(\mu|m)) \end{aligned} \tag{7}$$

Доказательство. Мы приведем набросок доказательств. Хотя некоторые аргументы могут показаться абстрактными, они становятся вполне элементарными на конечных или счётных пространствах.

1. Возьмем f константой в определении H (также следует из следующего пункта).
2. Функция $\mathcal{P}(E) \times \mathcal{P}(E) \ni (\mu, m) \mapsto \mu(f) - \log m(e^f) \in \mathbb{R}$ выпукла для каждой f, так как супремум выпуклых функций выпуклый.
3. Для простоты рассмотрим случай, когда E компактно. Тогда, поскольку $C(E)$ плотно в $L^1(\mu)$, все равно, брать ли супремум по $f \in C(E)$ в определении H. Но тогда отображение $\mathcal{P}(E) \times \mathcal{P}(E) \ni (\mu, m) \mapsto \mu(f) - \log m(e^f) \in \mathbb{R}$ непрерывно в метрике Леви — Прохорова, а супремум непрерывных функций полунепрерывен снизу. Если E локально компактно, можно заменить $C(E)$ на функции с компактным носителем. В общем случае мы можем заменить $C(E)$ на функции, которые равномерно непрерывны относительно фиксированной вполне ограниченной метрики на E (такая метрика всегда существует на польских пространствах).
4. Если существует множество A, такое что $\mu(A) > 0$, но $m(A) = 0$, то для $c > 0$ возьмем $f = c\mathbf{1}_A$. Получим

   $$H(\mu|m) \geq c\mu(A) - \log(e^c m(A) + e^0 m(A^c)) = c\mu(A)$$

   Поскольку это выполняется для любого $c > 0$, $H(\mu|m) = +\infty$. Если такого множества A не существует, по теореме Радона — Никодима, мы можем предположить, что $\mu = \varrho m$ для некоторого $\varrho \in L^1(m)$. Сначала предположим, что ϱ ограничена и отделена от 0. Тогда возьмем $f = \log \varrho + g$ для некоторой произвольной ограниченной измеримой g, чтобы получить

   $$\begin{aligned} H(\mu|m) &= \sup_g \int \log \varrho d\mu + \int g d\mu - \log \int e^{\log \varrho + g} dm \\ &= \int h(\varrho) dm + \sup_g \int g d\mu - \log \int e^g d\mu \end{aligned}$$

   Последний супремум неположителен по неравенству Йенсена, поэтому $\sup_g$ равен 0 (достигается при g константе). Затем нетрудно адаптировать аргумент для случая, когда $\log \varrho$ неограничен.
5. Непосредственно следует из свойств логарифма и цепного правила.
6. Возьмите $f = c\mathbf{1}_A$ и оптимизируйте по $c > 0$.
7. Это неравенство Пинскера, которое может быть доказано элементарными методами, но выходит за рамки данной заметки.

$\square$

Связь с классической энтропией

Примечание. На конечном пространстве E можно определить $\text{Ent}(\mu) := \sum_{x\in E}\mu_x \log \mu_x$. Заметим, что в записи относительной энтропии это есть не что иное, как

$$\text{Ent}(\mu) = H(\mu|m') - \log(|E|)$$

где m' — равномерная вероятность на E, то есть $m'_x = 1/|E|$ для всех $x \in E$. Из Уравнение 6, если $m_x = e^{-V(x)}/Z$ для некоторой $V\colon E \to \mathbb{R}\cup\{+\infty\}$ и $Z = \sum_x e^{-V(x)}$, следует

$$H(\mu|m) = \text{Ent}(\mu) + \int V d\mu + \log(Z/|E|)$$

В физической литературе $S(\mu) := -\text{Ent}(\mu)$ называется энтропией, $V(x)$ интерпретируется как $V(x) = \beta h(x)$, где $h(x)$ — энергия конфигурации состояния $x \in E$, $\beta = 1/(\kappa T)$, где κ — универсальная константа (*постоянная Больцмана*), а T — *температура*. $\int h\mu$ интерпретируется как *энергия* «состояния» μ. Таким образом, определяется *свободная энергия*

$$F(\mu) = \text{энергия - } \kappa T \text{ энтропия } = \int h d\mu - \kappa T S(\mu) = \tfrac{1}{\beta}H(\mu|m) - \tfrac{1}{\beta}\log(Z/|E|)$$

В частности, утверждение «$\mu \mapsto H(\mu|m)$ минимизируется при $\mu = m$» эквивалентно перефразируется как «мера $m = e^{-h/(\kappa T)}/Z$ минимизирует свободную энергию». К сожалению, различные обозначения и номенклатура сохраняются по сей день, и как общее правило:

- $H(\mu|m)$ используется в вероятностной литературе независимо от пространства.
- $D_{KL}(\mu||m)$ используется в литературе по информатике. Это совпадает с $H(\mu|m)$.
- $F(\mu)$ используется в физической литературе. Она отличается от $H(\mu|m)$ на некоторые константы, которые не существенны при фиксированных β или h (например, если мы рассматриваем их как функции от μ), но существенны, если мы рассматриваем $H(\mu|m)$ как функцию и от μ, и от m.

Большие уклонения

В этом разделе мы рассмотрим некоторые основные свойства концентрации последовательностей вероятностных мер. Как и выше, E — польское пространство, снабженное борелевской σ-алгеброй.

Определение 0.6. Функция $I\colon E \to (-\infty, \infty]$ называется **полунепрерывной снизу** (или **lsc**), если для всех $c \in \mathbb{R}$ множество $\{I \le c\}$ замкнуто.

I называется **коэрцитивной**, если $\{I \le c\}$ либо пусто, либо предкомпактно.

В частности, если I полунепрерывна снизу и коэрцитивна, то она достигает минимума на E[2].

Утверждение 0.4. *Пусть (μ_n) — последовательность вероятностных мер на E, а* $\mathbf{a} = (a_n)$ *— последовательность вещественных чисел с* $\lim_{n\to\infty} a_n = +\infty$. *Пусть $B_\varepsilon(x)$ — шар радиуса ε с центром в x. Определим*

$$\underline{I}(x) \equiv \underline{I}^{\mathbf{a}}(x) := -\lim_{\varepsilon\to 0}\overline{\lim_{n}}\frac{1}{a_n}\log\mu_n(B_\varepsilon(x)) \in [0,\infty]$$
$$\overline{I}(x) \equiv \overline{I}^{\mathbf{a}}(x) := -\lim_{\varepsilon\to 0}\underline{\lim_{n}}\frac{1}{a_n}\log\mu_n(B_\varepsilon(x)) \in [0,\infty]$$

[2]Этот простой для доказательства факт, известный как теорема Больцано — Вейерштрасса, обычно обсуждается на курсах матанализа, и строгость доказательства сыграла важную роль для вдохновения современной математики.

Тогда:

1. $\underline{I}$ и $\overline{I}$ *полунепрерывны снизу (lsc).*
2. $\underline{I}$ *является оптимальной (т.е. наибольшей) lsc функцией, а* $\overline{I}$ *является оптимальной (т.е. наименьшей) функцией, таких что выполняются следующие неравенства*

$$\mu_n(K) \leq \exp\left(-a_n \inf_{x\in K} \underline{I}(x) + o(a_n)\right), \qquad \text{для всех } K \subset E$$
$$\mu_n(O) \geq \exp\left(-a_n \inf_{x\in O} \underline{I}(x) + o(a_n)\right), \qquad \text{для всех } O \subset E$$

3. *Эквивалентно, они являются оптимальными (lsc) функциями, такими что для каждой* $f \in C_b(E)$

$$\mu_n(e^{a_n f}) \leq \exp\left(a_n \sup_x (f(x) - \underline{I}(x)) + o(a_n)\right)$$
$$\mu_n(e^{a_n f}) \geq \exp\left(a_n \sup_x (f(x) - \overline{I}(x)) + o(a_n)\right)$$

Примечание. Пусть $x \in E$. Для каждой последовательности $\nu_n \to \delta_x$ выполняется

$$\varliminf_n \frac{1}{a_n} H(\nu_n|\mu_n) \geq \underline{I}^{\mathbf{a}}(x)$$

Более того, существует последовательность $\nu_n \to \delta_x$, такая что

$$\varliminf_n \frac{1}{a_n} H(\nu_n|\mu_n) \leq \overline{I}^{\mathbf{a}}(x)$$

$\underline{I}^{\mathbf{a}}(x), \overline{I}^{\mathbf{a}}(x)$ являются оптимальными функциями, для которых выполняются эти два утверждения.

Большие уклонения следует сравнивать со слабой сходимостью Определение 0.4, в которой сходимость по существу эквивалентна $\mu_n(f) \to \mu(f)$ для $f \in C_{\mathrm{b}}(E)$, и аналогичные неравенства выполняются на открытых и замкнутых множествах. Неформально говоря, сходимость вероятностных мер соответствует случаю $a_n = 1$ из Утверждение 0.4.

Определение 0.7. Пусть (μ_n), (a_n) — как в Утверждение 0.4. Говорят, что (μ_n) удовлетворяет **принципу больших уклонений** со скоростью a_n и функцией уклонений $I\colon E \to [0, \infty]$, если $\underline{I} = \overline{I} =: I$. Принцип больших уклонений называется **нетривиальным**, если существует $x \in E$, такой что $I(x) \in (0, \infty)$.

Утверждение 0.5. *Пусть* (μ_n), (a_n) *— как в Утверждение 0.4. Существует подпоследовательность* n_k, *вдоль которой выполняется принцип больших уклонений.*

Неравенства

Здесь мы перечислим некоторые замечательные неравенства. Доказательства элементарны, за исключением неравенства Альсведе — Дэйкина на произведениях пространств.

Неравенства Маркова

Утверждение 0.6 (Неравенство Маркова). *Пусть X — вещественная случайная величина, тогда для $c > 0$*

$$\mathbb{P}[X \geq c] \leq \mathbb{E}[|X|]/c$$

Хотя это кажется грубым неравенством, мы можем применить его к любой неубывающей функции $\varphi\colon \mathbb{R} \to \mathbb{R}^+$, чтобы получить для любых X и $c \in \mathbb{R}$

$$\mathbb{P}[X \geq c] \leq \mathbb{P}[\varphi(X) \geq \varphi(c)] \leq \mathbb{E}[\varphi(X)]/\varphi(c)$$

С другой стороны, последнее тривиально становится равенством для $\varphi(x) = \mathbf{1}_{(\infty,c]}(x)$. Это влечет за собой следующее более сильное утверждение

Утверждение 0.7 (Равенство Маркова). *Для любой вещественной случайной величины и $c \in \mathbb{R}$ выполняется*

$$\mathbb{P}[X \geq c] = \inf_{\varphi} \mathbb{E}[\varphi(X)]/\varphi(c)$$

где инфимум берется по всем неубывающим $\varphi\colon \mathbb{R} \to [0, \infty)$ с $\varphi(c) > 0$.

В частности:

- *принимая $\varphi(x) = |x - \mathbb{E}[X]|$, мы получаем* ***неравенство Чебышёва***

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq c] \leq \mathrm{Var}[X]c^{-2}$$

- *принимая $\varphi(x) = e^{\lambda x}$ для $\lambda > 0$, мы получаем* ***неравенство Чернова***

$$\mathbb{P}[X \geq c] \leq \exp(-\psi(c))$$

где ψ задается формулой двойственности Лежандра $\psi(c) := \sup_{\lambda \geq 0} \lambda c - \log \mathbb{E}[e^{\lambda X}] \in [0, +\infty]$.

Неравенства Йенсена

Поскольку выпуклая функция является супремумом аффинных функций, меняя местами супремум и математическое ожидание, мы получаем знаменитое неравенство Йенсена, которое справедливо для довольно общих линейных пространств.

Утверждение 0.8 (Неравенство Йенсена). *Если f — выпуклая функция, X — случайная величина, такая что $f(X) \in L^1(\mathbb{P})$, то выполняется*

$$\mathbb{E}[f(X)|\mathcal{G}] \geq f\left(\mathbb{E}[X|\mathcal{G}]\right)$$

Неравенства Гёльдера

Утверждение 0.9 (Неравенство Гёльдера). *Если $p_1, \ldots, p_n, q \in [1, \infty]$ таковы, что $\sum_i 1/p_i \leq 1/q$, то*

$$\mathbb{E}[|X_1 \cdots X_n|^q]^{1/q} \leq \mathbb{E}[|X_1|^{p_1}]^{1/p_1} \cdots \mathbb{E}[|X_n|^{p_n}]^{1/p_n}$$

Лемма Кошена — Стоуна

Это аналог леммы Бореля — Кантелли

Утверждение 0.10 (Лемма Кошена — Стоуна). *Пусть (A_n) — последовательность событий, такая что*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$$

Тогда

$$\mathbb{P}(\limsup_{n\to\infty} A_n) \geq \limsup_{k\to\infty} \frac{\left(\sum_{n=1}^{k} \mathbb{P}(A_n)\right)^2}{\sum_{1\leq m,n\leq k} \mathbb{P}(A_m \cap A_n)} \tag{8}$$

В частности, если (A_n) попарно независимы (или A_n не зависит от A_m для всех, кроме конечного числа m), то при условии Уравнение 8 $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.

Корреляционные неравенства

В этом разделе мы вводим два класса нетривиальных корреляционных неравенств: неравенства класса GKS/Жинибра и неравенства класса FKG.

Неравенство Жинибра

Определение 0.8 (Выпуклый конус). Подмножество C вещественного векторного пространства называется (тупым) **выпуклым конусом**, если оно замкнуто относительно линейных комбинаций с неотрицательными коэффициентами: если $u, v \in C$, то $\alpha u + \beta v \in C$ для $\alpha, \beta \geq 0$.

Наименьший выпуклый конус, содержащий подмножество A векторного пространства, называется **выпуклым конусом, порожденным** A (это корректно определено как пересечение всех выпуклых конусов, содержащих A).

Определение 0.9. Пусть $A \subset L^1(\mathbb{P})$ — набор интегрируемых случайных величин. Мы говорим, что A удовлетворяет **условию Жинибра**, если для каждого $N \geq 1$, $X_1, \dots, X_N \in A$ и $\epsilon_1, \dots, \epsilon_N \in \{-1, +1\}$

$$\int \prod_{i=1}^{N} (X_i(\omega) + \epsilon_i X_i(\omega'))\, \mathbb{P}(d\omega)\mathbb{P}(d\omega') \geq 0$$

Другими словами, если $(Y_1, \dots, Y_N)$ — независимая копия $(X_1, \dots, X_N)$, то $\mathbb{E}[(X_1 \pm Y_1) \cdots (X_N \pm Y_N)] \geq 0$, независимо от знаков $\pm$ в каждом множителе.

Теорема 0.1 (Неравенство Жинибра). *Пусть $A \subset L^1(\mathbb{P})$ — набор интегрируемых случайных величин, удовлетворяющих условию Жинибра. Если X, Y, H — случайные величины из выпуклого конуса, порожденного A, и $e^{-H}, Xe^{-H}, Ye^{-H} \in L^1(\mathbb{P})$, то*

$$\mathbb{E}\left[XYe^{-H}\right] \mathbb{E}\left[e^{-H}\right] \geq \mathbb{E}\left[Xe^{-H}\right] \mathbb{E}\left[Ye^{-H}\right]$$

Общие неравенства AD и FKG

Чтобы корректно определить общую версию неравенств AD, Холли и FKG, мы сначала напомним понятие дистрибутивной решётки.

Определение 0.10 (Измеримая дистрибутивная решётка)**.** Отношение частичного порядка $\preccurlyeq$, определенное на измеримом пространстве $(\Omega, \mathcal{G})$, называется **измеримым**, если множество $\{(x,y)\ :\ x \preccurlyeq y\}$ измеримо.

Множество Ω, снабженное отношением частичного порядка $\preccurlyeq$, является **дистрибутивной решёткой**, если для любых элементов $x, y, z \in \Omega$:

a. Существуют единственная точная нижняя грань $x \wedge y$ и единственная точная верхняя грань $x \vee y$ (**свойство решётки**).
b. Операции **дистрибутивны**: $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$.

Измеримая дистрибутивная решётка — это тройка $(\Omega, \mathcal{F}, \preccurlyeq)$, где $(\Omega, \mathcal{F})$ — измеримое пространство, а $\preccurlyeq$ — измеримый частичный порядок, определяющий структуру дистрибутивной решётки на Ω.

Определение 0.11 (Корреляционные неравенства)**.** Пусть μ — σ-конечная мера на измеримой дистрибутивной решётке $(\Omega, \mathcal{F}, \preccurlyeq)$.

- Мы говорим, что μ удовлетворяет **неравенству Альсведе — Дэйкина**, если для всех измеримых $f_1, f_2, f_3, f_4 \colon \Omega \to [0, \infty]$, таких что для $x, y \in \Omega$

  $$f_1(x \vee y) f_2(x \wedge y) \geq f_3(x) f_4(y) \qquad x, y \in \Omega$$

  выполняется

  $$\mu(f_1)\mu(f_2) \geq \mu(f_3)\mu(f_4) \tag{9}$$

- Мы говорим, что μ удовлетворяет **неравенству Холли**, если для всех измеримых функций $h, g_1, g_2 \colon \Omega \to [0, \infty]$, таких что h неубывающая и

  $$g_1(x \vee y) g_2(x \wedge y) \geq g_1(x) g_2(y) \qquad x, y \in \Omega$$

  выполняется

  $$\mu(h\, g_1)\mu(g_2) \geq \mu(g_1)\mu(h g_2) \tag{10}$$

- Если μ — вероятность, мы говорим, что μ удовлетворяет **неравенству FKG**, если для всех измеримых неубывающих функций $f, g \colon \Omega \to [0, \infty]$

  $$\mu(fg) \geq \mu(f)\mu(g)$$

Утверждение 0.11. *Если σ-конечная мера μ удовлетворяет неравенству Альсведе — Дэйкина, то она удовлетворяет неравенству Холли.*

Если вероятность μ удовлетворяет неравенству Холли, то она удовлетворяет неравенству FKG.

Доказательство. Для первого утверждения возьмем в Уравнение 9 $f_1 = hg_1$, $f_2 = g_2$, $f_3 = g_1$, $f_4 = hg_2$. Легко видеть, что они удовлетворяют условиям неравенства Альсведе — Дэйкина, так как $h(x \vee y) \geq h(y)$. Но для такого выбора четырех функций неравенство Альсведе — Дэйкина сводится к неравенству Холли.

Для второго утверждения мы можем предположить, что $f \in L^1(\mu)$, с точностью до простой аппроксимации. Тогда возьмем в Уравнение 10 $g_1 = f$, $g_2 = 1$, $h = g$. $\square$

Особым классом измеримых дистрибутивных решёток являются решётки произведений. Предположим, что для t из некоторого произвольного множества индексов T, $(\Omega_t, \mathcal{F}_t, \preccurlyeq_t)$ является измеримой дистрибутивной решёткой, и предположим, что $\preccurlyeq_t$ является отношением **линейного** порядка. Тогда пространство-произведение $\Omega = \prod_{t\in T} \Omega_t$ естественным образом снабжается отношением частичного порядка: $\omega \preccurlyeq \omega'$ тогда и только тогда, когда $\omega_t \preccurlyeq_t \omega'_t$ для всех $t \in T$. В этом случае мы говорим, что Ω является измеримой дистрибутивной решёткой **произведения**.

Основное утверждение следующей теоремы доказано в (Batty и Bollmann 1980 г.).

Теорема 0.2 (Общая теорема Альсведе — Дэйкина)**.** *Любая мера произведения на дистрибутивной решётке произведения удовлетворяет неравенству Альсведе — Дэйкина.*

В частности, поскольку любую конечную дистрибутивную решётку можно рассматривать как подрешётку (конечной) дистрибутивной решётки произведения, мы имеем, что считающая мера на конечной дистрибутивной решётке удовлетворяет неравенству Альсведе — Дэйкина.

Batty, CJK, и HW Bollmann. 1980 г. «Generalised Holley-Preston inequalities on measure spaces and their products». *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 53 (2): 157–73.