

# Concentration Inequalities

Emin Karayel and Yong Kiam Tan\*

April 20, 2024

## Abstract

Concentration inequalities provide bounds on how a random variable (or a sum/composition of random variables) deviate from their expectation, usually based on moments/independence of the variables.

The most important concentration inequalities (the Markov, Chebyshev, and Hoelder inequalities and the Chernoff bounds) are already part of HOL-Probability. This entry collects more advanced results, such as Bennett's/Bernstein's Inequality, Bienaymé's Identity, Cantelli's Inequality, the Efron-Stein Inequality, McDiarmid's Inequality, and the Paley-Zygmund Inequality.

## Contents

<b>1</b>	<b>Preliminary results</b>	<b>1</b>
<b>2</b>	<b>Bennett's Inequality</b>	<b>4</b>
<b>3</b>	<b>Bienaymé's identity</b>	<b>6</b>
<b>4</b>	<b>Cantelli's Inequality</b>	<b>9</b>
<b>5</b>	<b>Efron-Stein Inequality</b>	<b>10</b>
<b>6</b>	<b>McDiarmid's inequality</b>	<b>10</b>
<b>7</b>	<b>Paley-Zygmund Inequality</b>	<b>14</b>

## 1 Preliminary results

```
theory Concentration-Inequalities-Preliminary
  imports Lp.Lp
begin
```

Version of Cauchy-Schwartz for the Lebesgue integral:

---

\*The authors contributed equally to this work.

**lemma** *cauchy-schwartz*:

**fixes**  $f\ g :: - \Rightarrow \text{real}$   
**assumes**  $f \in \text{borel-measurable } M\ g \in \text{borel-measurable } M$   
**assumes**  $\text{integrable } M (\lambda x. (f\ x)^{\wedge 2})\ \text{integrable } M (\lambda x. (g\ x)^{\wedge 2})$   
**shows**  $\text{integrable } M (\lambda x. f\ x * g\ x)$  (**is** ?A)  
 $(\int x. f\ x * g\ x\ \partial M) \leq (\int x. (f\ x)^{\wedge 2}\ \partial M)\ \text{powr } (1/2) * (\int x. (g\ x)^{\wedge 2}\ \partial M)\ \text{powr } (1/2)$   
(**is** ?L  $\leq$  ?R)  
{proof}

Generalization of *prob-space.indep-vars-iff-distr-eq-PiM'*:

**lemma** (**in** *prob-space*) *indep-vars-iff-distr-eq-PiM''*:

**fixes**  $I :: 'i\ \text{set}$  **and**  $X :: 'i \Rightarrow 'a \Rightarrow 'b$   
**assumes**  $rv: \bigwedge i. i \in I \Longrightarrow \text{random-variable } (M'\ i)\ (X\ i)$   
**shows**  $\text{indep-vars } M'\ X\ I \longleftrightarrow$   
 $\text{distr } M\ (\Pi_M\ i \in I. M'\ i)\ (\lambda x. \lambda i \in I. X\ i\ x) = (\Pi_M\ i \in I. \text{distr } M\ (M'\ i)\ (X\ i))$   
{proof}

**lemma** *proj-indep*:

**assumes**  $\bigwedge i. i \in I \Longrightarrow \text{prob-space } (M\ i)$   
**shows**  $\text{prob-space.indep-vars } (PiM\ I\ M)\ M\ (\lambda i\ \omega. \omega\ i)\ I$   
{proof}

**lemma** *forall-Pi-to-PiE*:

**assumes**  $\bigwedge x. P\ x = P\ (\text{restrict } x\ I)$   
**shows**  $(\forall x \in Pi\ I\ A. P\ x) = (\forall x \in PiE\ I\ A. P\ x)$   
{proof}

**lemma** *PiE-reindex*:

**assumes**  $\text{inj-on } f\ I$   
**shows**  $PiE\ I\ (A \circ f) = (\lambda a. \text{restrict } (a \circ f)\ I)\ 'PiE\ (f\ 'I)\ A$  (**is** ?lhs = ?g ' ?rhs)  
{proof}

**context** *prob-space*

**begin**

**lemma** *indep-sets-reindex*:

**assumes**  $\text{inj-on } f\ I$   
**shows**  $\text{indep-sets } A\ (f\ 'I) = \text{indep-sets } (\lambda i. A\ (f\ i))\ I$   
{proof}

**lemma** *indep-vars-reindex*:

**assumes**  $\text{inj-on } f\ I$   
**assumes**  $\text{indep-vars } M'\ X'\ (f\ 'I)$   
**shows**  $\text{indep-vars } (M' \circ f)\ (\lambda k\ \omega. X'\ (f\ k)\ \omega)\ I$   
{proof}

**lemma** *indep-vars-cong-AE*:  
**assumes**  $AE\ x\ in\ M. (\forall i \in I. X' i\ x = Y' i\ x)$   
**assumes**  $indep\ vars\ M' X' I$   
**assumes**  $\bigwedge i. i \in I \implies random\ variable\ (M' i)\ (Y' i)$   
**shows**  $indep\ vars\ M' Y' I$   
 $\langle proof \rangle$

**end**

Integrability of bounded functions on finite measure spaces:

**lemma** *bounded-const*:  $bounded\ ((\lambda x. (c::real))\ 'T)$   
 $\langle proof \rangle$

**lemma** *bounded-exp*:  
**fixes**  $f :: 'a \Rightarrow real$   
**assumes**  $bounded\ ((\lambda x. f\ x)\ 'T)$   
**shows**  $bounded\ ((\lambda x. exp\ (f\ x))\ 'T)$   
 $\langle proof \rangle$

**lemma** *bounded-mult-comp*:  
**fixes**  $f :: 'a \Rightarrow real$   
**assumes**  $bounded\ (f\ 'T)\ bounded\ (g\ 'T)$   
**shows**  $bounded\ ((\lambda x. (f\ x) * (g\ x))\ 'T)$   
 $\langle proof \rangle$

**lemma** *bounded-sum*:  
**fixes**  $f :: 'i \Rightarrow 'a \Rightarrow real$   
**assumes**  $finite\ I$   
**assumes**  $\bigwedge i. i \in I \implies bounded\ (f\ i\ 'T)$   
**shows**  $bounded\ ((\lambda x. (\sum i \in I. f\ i\ x))\ 'T)$   
 $\langle proof \rangle$

**lemma** *bounded-pow*:  
**fixes**  $f :: 'a \Rightarrow real$   
**assumes**  $bounded\ ((\lambda x. f\ x)\ 'T)$   
**shows**  $bounded\ ((\lambda x. (f\ x)^n)\ 'T)$   
 $\langle proof \rangle$

**lemma** *bounded-sum-list*:  
**fixes**  $f :: 'i \Rightarrow 'a \Rightarrow real$   
**assumes**  $\bigwedge y. y \in set\ ys \implies bounded\ (f\ y\ 'T)$   
**shows**  $bounded\ ((\lambda x. (\sum y \leftarrow ys. f\ y\ x))\ 'T)$   
 $\langle proof \rangle$

**lemma** (*in finite-measure*) *bounded-int*:  
**fixes**  $f :: 'i \Rightarrow 'a \Rightarrow real$   
**assumes**  $bounded\ ((\lambda x. f\ (fst\ x)\ (snd\ x))\ '(T \times space\ M))$   
**shows**  $bounded\ ((\lambda x. (\int \omega. (f\ x\ \omega)\ \partial M))\ 'T)$   
 $\langle proof \rangle$

**lemmas** *bounded-intros* =  
*bounded-minus-comp bounded-plus-comp bounded-mult-comp bounded-sum*  
*finite-measure.bounded-int*  
*bounded-const bounded-exp bounded-pow bounded-sum-list*

**lemma** (*in prob-space*) *integrable-bounded*:  
**fixes**  $f :: - \Rightarrow ('b :: \{\text{banach,second-countable-topology}\})$   
**assumes** *bounded* ( $f \text{ ' } \text{space } M$ )  
**assumes**  $f \in M \rightarrow_M \text{borel}$   
**shows** *integrable*  $M f$   
 $\langle \text{proof} \rangle$

**lemma** *integrable-bounded-pmf*:  
**fixes**  $f :: - \Rightarrow ('b :: \{\text{banach,second-countable-topology}\})$   
**assumes** *bounded* ( $f \text{ ' } \text{set-pmf } M$ )  
**shows** *integrable* (*measure-pmf*  $M$ )  $f$   
 $\langle \text{proof} \rangle$

**end**

## 2 Bennett's Inequality

In this section we verify Bennett's inequality [1] and a (weak) version of Bernstein's inequality as a corollary. Both inequalities give concentration bounds for sums of independent random variables. The statement and proofs follow a summary paper by Boucheron et al. [2].

**theory** *Bennett-Inequality*  
**imports** *Concentration-Inequalities-Preliminary*  
**begin**

**context** *prob-space*  
**begin**

**lemma** *indep-vars-Chernoff-ineq-ge*:  
**assumes**  $I: \text{finite } I$   
**assumes** *ind: indep-vars* ( $\lambda \text{ -. borel}$ )  $X I$   
**assumes** *sge*:  $s \geq 0$   
**assumes** *int*:  $\bigwedge i. i \in I \implies \text{integrable } M (\lambda x. \exp (s * X i x))$   
**shows** *prob*  $\{x \in \text{space } M. (\sum i \in I. X i x - \text{expectation } (X i)) \geq t\} \leq$   
 $\exp (-s*t) * (\prod i \in I. \text{expectation } (\lambda x. \exp(s * (X i x - \text{expectation } (X i)))))$   
 $\langle \text{proof} \rangle$

**definition** *bennett-h::real*  $\Rightarrow \text{real}$

where  $\text{bennett-h } u = (1 + u) * \ln (1 + u) - u$

**lemma** *exp-sub-two-terms-eq*:

**fixes**  $x :: \text{real}$

**shows**  $\exp x - x - 1 = (\sum n. x^{(n+2)} / \text{fact } (n+2))$   
 $\text{summable } (\lambda n. x^{(n+2)} / \text{fact } (n+2))$

*<proof>*

**lemma** *psi-mono*:

**defines**  $f \equiv (\lambda x. (\exp x - x - 1) - x^2 / 2)$

**assumes**  $xy: a \leq (b::\text{real})$

**shows**  $f a \leq f b$

*<proof>*

**lemma** *psi-inequality*:

**assumes**  $le: x \leq (y::\text{real}) y \geq 0$

**shows**  $y^2 * (\exp x - x - 1) \leq x^2 * (\exp y - y - 1)$

*<proof>*

**lemma** *bennett-inequality-1*:

**assumes**  $I: \text{finite } I$

**assumes**  $ind: \text{indep-vars } (\lambda -. \text{borel}) X I$

**assumes**  $intsq: \bigwedge i. i \in I \implies \text{integrable } M (\lambda x. (X i x)^2)$

**assumes**  $bnd: \bigwedge i. i \in I \implies \text{AE } x \text{ in } M. X i x \leq 1$

**assumes**  $t: t \geq 0$

**defines**  $V \equiv (\sum i \in I. \text{expectation}(\lambda x. X i x^2))$

**shows**  $\text{prob } \{x \in \text{space } M. (\sum i \in I. X i x - \text{expectation } (X i)) \geq t\} \leq$   
 $\exp (-V * \text{bennett-h } (t / V))$

*<proof>*

**lemma** *real-AE-le-sum*:

**assumes**  $\bigwedge i. i \in I \implies \text{AE } x \text{ in } M. f i x \leq (g i x::\text{real})$

**shows**  $\text{AE } x \text{ in } M. (\sum i \in I. f i x) \leq (\sum i \in I. g i x)$

*<proof>*

**lemma** *real-AE-eq-sum*:

**assumes**  $\bigwedge i. i \in I \implies \text{AE } x \text{ in } M. f i x = (g i x::\text{real})$

**shows**  $\text{AE } x \text{ in } M. (\sum i \in I. f i x) = (\sum i \in I. g i x)$

*<proof>*

**theorem** *bennett-inequality*:

**assumes**  $I: \text{finite } I$

**assumes**  $ind: \text{indep-vars } (\lambda -. \text{borel}) X I$

**assumes**  $intsq: \bigwedge i. i \in I \implies \text{integrable } M (\lambda x. (X i x)^2)$

**assumes**  $bnd: \bigwedge i. i \in I \implies \text{AE } x \text{ in } M. X i x \leq B$

```

assumes  $t: t \geq 0$ 
assumes  $B: B > 0$ 
defines  $V \equiv (\sum i \in I. \text{expectation } (\lambda x. X i x^2))$ 
shows  $\text{prob } \{x \in \text{space } M. (\sum i \in I. X i x - \text{expectation } (X i)) \geq t\} \leq$ 
 $\text{exp } (- V / B^2 * \text{bennett-h } (t * B / V))$ 
<proof>

```

```

lemma bennett-h-bernstein-bound:
assumes  $x \geq 0$ 
shows  $\text{bennett-h } x \geq x^2 / (2 * (1 + x / 3))$ 
<proof>

```

```

lemma sum-sq-exp-eq-zero-imp-zero:
assumes  $\text{finite } I \ i \in I$ 
assumes  $\text{intsq: integrable } M \ (\lambda x. (X i x)^2)$ 
assumes  $(\sum i \in I. \text{expectation } (\lambda x. X i x^2)) = 0$ 
shows  $\text{AE } x \text{ in } M. X i x = (0::\text{real})$ 
<proof>

```

```

corollary bernstein-inequality:
assumes  $I: \text{finite } I$ 
assumes  $\text{ind: indep-vars } (\lambda -. \text{borel}) \ X \ I$ 
assumes  $\text{intsq: } \bigwedge i. i \in I \implies \text{integrable } M \ (\lambda x. (X i x)^2)$ 
assumes  $\text{bnd: } \bigwedge i. i \in I \implies \text{AE } x \text{ in } M. X i x \leq B$ 
assumes  $t: t \geq 0$ 
assumes  $B: B > 0$ 
defines  $V \equiv (\sum i \in I. \text{expectation } (\lambda x. X i x^2))$ 
shows  $\text{prob } \{x \in \text{space } M. (\sum i \in I. X i x - \text{expectation } (X i)) \geq t\} \leq$ 
 $\text{exp } (- (t^2 / (2 * (V + t * B / 3))))$ 
<proof>

```

end

end

### 3 Bienaymé's identity

Bienaymé's identity [5, §17] can be used to deduce the variance of a sum of random variables, if their co-variance is known. A common use-case of the identity is the computation of the variance of the mean of pair-wise independent variables.

```

theory Bienaymes-Identity
imports Concentration-Inequalities-Preliminary
begin

```

**context** *prob-space*  
**begin**

**lemma** *variance-divide*:

**fixes**  $f :: 'a \Rightarrow \text{real}$   
**assumes** *integrable*  $M f$   
**shows**  $\text{variance } (\lambda \omega. f \ \omega / r) = \text{variance } f / r^2$   
*<proof>*

**definition** *covariance* **where**

$\text{covariance } f g = \text{expectation } (\lambda \omega. (f \ \omega - \text{expectation } f) * (g \ \omega - \text{expectation } g))$

**lemma** *covariance-eq*:

**fixes**  $f :: 'a \Rightarrow \text{real}$   
**assumes**  $f \in \text{borel-measurable } M$   $g \in \text{borel-measurable } M$   
**assumes** *integrable*  $M (\lambda \omega. f \ \omega^2)$  *integrable*  $M (\lambda \omega. g \ \omega^2)$   
**shows**  $\text{covariance } f g = \text{expectation } (\lambda \omega. f \ \omega * g \ \omega) - \text{expectation } f$   
 $* \text{expectation } g$   
*<proof>*

**lemma** *covar-integrable*:

**fixes**  $f g :: 'a \Rightarrow \text{real}$   
**assumes**  $f \in \text{borel-measurable } M$   $g \in \text{borel-measurable } M$   
**assumes** *integrable*  $M (\lambda \omega. f \ \omega^2)$  *integrable*  $M (\lambda \omega. g \ \omega^2)$   
**shows** *integrable*  $M (\lambda \omega. (f \ \omega - \text{expectation } f) * (g \ \omega - \text{expectation } g))$   
*<proof>*

**lemma** *sum-square-int*:

**fixes**  $f :: 'b \Rightarrow 'a \Rightarrow \text{real}$   
**assumes** *finite*  $I$   
**assumes**  $\bigwedge i. i \in I \implies f \ i \in \text{borel-measurable } M$   
**assumes**  $\bigwedge i. i \in I \implies \text{integrable } M (\lambda \omega. f \ i \ \omega^2)$   
**shows** *integrable*  $M (\lambda \omega. (\sum i \in I. f \ i \ \omega)^2)$   
*<proof>*

**theorem** *bienaymes-identity*:

**fixes**  $f :: 'b \Rightarrow 'a \Rightarrow \text{real}$   
**assumes** *finite*  $I$   
**assumes**  $\bigwedge i. i \in I \implies f \ i \in \text{borel-measurable } M$   
**assumes**  $\bigwedge i. i \in I \implies \text{integrable } M (\lambda \omega. f \ i \ \omega^2)$   
**shows**  
 $\text{variance } (\lambda \omega. (\sum i \in I. f \ i \ \omega)) = (\sum i \in I. (\sum j \in I. \text{covariance } (f \ i) (f \ j)))$   
*<proof>*

**lemma** *covar-self-eq*:

**fixes**  $f :: 'a \Rightarrow \text{real}$

**shows**  $\text{covariance } f f = \text{variance } f$   
*<proof>*

**lemma** *covar-indep-eq-zero*:  
**fixes**  $f g :: 'a \Rightarrow \text{real}$   
**assumes** *integrable*  $M f$   
**assumes** *integrable*  $M g$   
**assumes** *indep-var borel*  $f$  *borel*  $g$   
**shows**  $\text{covariance } f g = 0$   
*<proof>*

**lemma** *bienaymes-identity-2*:  
**fixes**  $f :: 'b \Rightarrow 'a \Rightarrow \text{real}$   
**assumes** *finite*  $I$   
**assumes**  $\bigwedge i. i \in I \Rightarrow f i \in \text{borel-measurable } M$   
**assumes**  $\bigwedge i. i \in I \Rightarrow \text{integrable } M (\lambda \omega. f i \omega^2)$   
**shows**  $\text{variance } (\lambda \omega. (\sum i \in I. f i \omega)) =$   
 $(\sum i \in I. \text{variance } (f i)) + (\sum i \in I. \sum j \in I - \{i\}. \text{covariance}$   
 $(f i) (f j))$   
*<proof>*

**theorem** *bienaymes-identity-pairwise-indep*:  
**fixes**  $f :: 'b \Rightarrow 'a \Rightarrow \text{real}$   
**assumes** *finite*  $I$   
**assumes**  $\bigwedge i. i \in I \Rightarrow f i \in \text{borel-measurable } M$   
**assumes**  $\bigwedge i. i \in I \Rightarrow \text{integrable } M (\lambda \omega. f i \omega^2)$   
**assumes**  $\bigwedge i j. i \in I \Rightarrow j \in I \Rightarrow i \neq j \Rightarrow \text{indep-var borel } (f i)$   
*borel*  $(f j)$   
**shows**  $\text{variance } (\lambda \omega. (\sum i \in I. f i \omega)) = (\sum i \in I. \text{variance } (f i))$   
*<proof>*

**lemma** *indep-var-from-indep-vars*:  
**assumes**  $i \neq j$   
**assumes** *indep-vars*  $(\lambda -. M') f \{i, j\}$   
**shows** *indep-var*  $M' (f i) M' (f j)$   
*<proof>*

**lemma** *bienaymes-identity-pairwise-indep-2*:  
**fixes**  $f :: 'b \Rightarrow 'a \Rightarrow \text{real}$   
**assumes** *finite*  $I$   
**assumes**  $\bigwedge i. i \in I \Rightarrow f i \in \text{borel-measurable } M$   
**assumes**  $\bigwedge i. i \in I \Rightarrow \text{integrable } M (\lambda \omega. f i \omega^2)$   
**assumes**  $\bigwedge J. J \subseteq I \Rightarrow \text{card } J = 2 \Rightarrow \text{indep-vars } (\lambda -. \text{borel}) f J$   
**shows**  $\text{variance } (\lambda \omega. (\sum i \in I. f i \omega)) = (\sum i \in I. \text{variance } (f i))$   
*<proof>*

**lemma** *bienaymes-identity-full-indep*:  
**fixes**  $f :: 'b \Rightarrow 'a \Rightarrow \text{real}$   
**assumes** *finite*  $I$



```

assumes  $\bigwedge i. i \in I \implies f\ i \in \text{borel-measurable } M$ 
assumes  $\bigwedge i. i \in I \implies \text{integrable } M (\lambda\omega. f\ i\ \omega^{\wedge 2})$ 
assumes indep-vars  $(\lambda -. \text{borel})\ f\ I$ 
shows variance  $(\lambda\omega. (\sum i \in I. f\ i\ \omega)) = (\sum i \in I. \text{variance } (f\ i))$ 
<proof>

```

**end**

**end**

## 4 Cantelli's Inequality

Cantelli's inequality [3] is an improvement of Chebyshev's inequality for one-sided tail bounds.

**theory** *Cantelli-Inequality*

**imports** *HOL-Probability.Probability*

**begin**

**context** *prob-space*

**begin**

**lemma** *cantelli-arith:*

**assumes**  $a > (0::\text{real})$

**shows**  $(V + (V / a)^{\wedge 2}) / (a + (V / a)^{\wedge 2})^{\wedge 2} = V / (a^{\wedge 2} + V)$  (**is**  
 $?L = ?R$ )

<proof>

**theorem** *cantelli-inequality:*

**assumes** [*measurable*]: *random-variable borel*  $Z$

**assumes** *intZsq*: *integrable*  $M (\lambda z. Z\ z^{\wedge 2})$

**assumes**  $a: a > 0$

**shows** *prob*  $\{z \in \text{space } M. Z\ z - \text{expectation } Z \geq a\} \leq$   
 $\text{variance } Z / (a^{\wedge 2} + \text{variance } Z)$

<proof>

**corollary** *cantelli-inequality-neg:*

**assumes** [*measurable*]: *random-variable borel*  $Z$

**assumes** *intZsq*: *integrable*  $M (\lambda z. Z\ z^{\wedge 2})$

**assumes**  $a: a > 0$

**shows** *prob*  $\{z \in \text{space } M. Z\ z - \text{expectation } Z \leq -a\} \leq$   
 $\text{variance } Z / (a^{\wedge 2} + \text{variance } Z)$

<proof>

**end**

**end**

## 5 Efron-Stein Inequality

In this section we verify the Efron-Stein inequality. The verified theorem is stated as Efron-Stein inequality for non-symmetric functions by Steele [8]. However most textbooks refer to this version as “the Efron-Stein inequality”. The original result that was shown by Efron and Stein is a tail bound for the variance of a symmetric functions of i.i.d. random variables [4].

**theory** *Efron-Stein-Inequality*

**imports** *Concentration-Inequalities-Preliminary*

**begin**

**theorem** *efron-stein-inequality-distr:*

**fixes**  $f :: - \Rightarrow \text{real}$

**assumes** *finite I*

**assumes**  $\bigwedge i. i \in I \implies \text{prob-space } (M\ i)$

**assumes** *integrable (PiM I M) ( $\lambda x. f\ x^2$ ) and f-meas:  $f \in \text{borel-measurable } (PiM\ I\ M)$*

**shows** *prob-space.variance (PiM I M)  $f \leq$*

*$(\sum_{i \in I}. (\int x. (f(\lambda j. x(j, \text{False})) - f(\lambda j. x(j, j=i)))^2 \partial PiM (I \times UNIV) (M \circ fst))) / 2$*

*(is ?L  $\leq$  ?R)*

*<proof>*

**theorem** *(in prob-space) efron-stein-inequality-classic:*

**fixes**  $f :: - \Rightarrow \text{real}$

**assumes** *finite I*

**assumes** *indep-vars (M' o fst) X (I  $\times$  (UNIV :: bool set))*

**assumes**  *$f \in \text{borel-measurable } (PiM\ I\ M')$*

**assumes** *integrable M ( $\lambda \omega. f(\lambda i \in I. X(i, \text{False})\ \omega)^2$ )*

**assumes**  $\bigwedge i. i \in I \implies \text{distr } M (M' i) (X(i, \text{True})) = \text{distr } M (M' i) (X(i, \text{False}))$

**shows** *variance ( $\lambda \omega. f(\lambda i \in I. X(i, \text{False})\ \omega) \leq$*

*$(\sum_{j \in I}. \text{expectation } (\lambda \omega. (f(\lambda i \in I. X(i, \text{False})\ \omega) - f(\lambda i \in I. X(i, i=j)\ \omega))^2) / 2$*

*(is ?L  $\leq$  ?R)*

*<proof>*

**end**

## 6 McDiarmid’s inequality

In this section we verify McDiarmid’s inequality [6, Lemma 1.2]. In the source and also further sources sometimes refer to the result as the “independent bounded differences” inequality.

**theory** *McDiarmid-Inequality*

**imports** *Concentration-Inequalities-Preliminary*

**begin**

**lemma** *Collect-restr-cong*:

**assumes**  $A = B$   
**assumes**  $\bigwedge x. x \in A \implies P x = Q x$   
**shows**  $\{x \in A. P x\} = \{x \in B. Q x\}$   
*<proof>*

**lemma** *ineq-chain*:

**fixes**  $h :: nat \Rightarrow real$   
**assumes**  $\bigwedge i. i < n \implies h (i+1) \leq h i$   
**shows**  $h n \leq h 0$   
*<proof>*

**lemma** *restrict-subset-eq*:

**assumes**  $A \subseteq B$   
**assumes**  $restrict f B = restrict g B$   
**shows**  $restrict f A = restrict g A$   
*<proof>*

Bochner Integral version of Hoeffding's Lemma using *interval-bounded-random-variable.Hoeffding*

**lemma** (**in** *prob-space*) *Hoeffdings-lemma-bochner*:

**assumes**  $l > 0$  **and**  $E0$ : *expectation*  $f = 0$   
**assumes** *random-variable* *borel*  $f$   
**assumes**  $AE x$  *in*  $M$ .  $f x \in \{a..b::real\}$   
**shows** *expectation*  $(\lambda x. exp (l * f x)) \leq exp (l^2 * (b - a)^2 / 8)$  (**is**  $?L \leq ?R$ )  
*<proof>*

**lemma** (**in** *prob-space*) *Hoeffdings-lemma-bochner-2*:

**assumes**  $l > 0$  **and**  $E0$ : *expectation*  $f = 0$   
**assumes** *random-variable* *borel*  $f$   
**assumes**  $\bigwedge x y. \{x,y\} \subseteq space M \implies |f x - f y| \leq (c::real)$   
**shows** *expectation*  $(\lambda x. exp (l * f x)) \leq exp (l^2 * c^2 / 8)$  (**is**  $?L \leq ?R$ )  
*<proof>*

**lemma** (**in** *prob-space*) *Hoeffdings-lemma-bochner-3*:

**assumes** *expectation*  $f = 0$   
**assumes** *random-variable* *borel*  $f$   
**assumes**  $\bigwedge x y. \{x,y\} \subseteq space M \implies |f x - f y| \leq (c::real)$   
**shows** *expectation*  $(\lambda x. exp (l * f x)) \leq exp (l^2 * c^2 / 8)$  (**is**  $?L \leq ?R$ )  
*<proof>*

Version of *product-sigma-finite.product-integral-singleton* without the condition that  $M i$  has to be sigma finite for all  $i$ :

**lemma** *product-integral-singleton*:

**fixes**  $f :: - \Rightarrow -::\{banach, second-countable-topology\}$

**assumes** *sigma-finite-measure* ( $M\ i$ )  
**assumes**  $f \in \text{borel-measurable}$  ( $M\ i$ )  
**shows**  $(\int x. f\ x\ \partial(\text{PiM}\ \{i\}\ M)) = (\int x. f\ x\ \partial(M\ i))$  (**is**  $?L = ?R$ )  
 <proof>

Version of *product-sigma-finite.product-integral-fold* without the condition that  $M\ i$  has to be sigma finite for all  $i$ :

**lemma** *product-integral-fold*:

**fixes**  $f :: - \Rightarrow -::\{\text{banach}, \text{second-countable-topology}\}$   
**assumes**  $\bigwedge i. i \in I \cup J \implies \text{sigma-finite-measure}$  ( $M\ i$ )  
**assumes**  $I \cap J = \{\}$   
**assumes** *finite*  $I$   
**assumes** *finite*  $J$   
**assumes** *integrable* ( $\text{PiM}\ (I \cup J)\ M$ )  $f$   
**shows**  $(\int x. f\ x\ \partial\text{PiM}\ (I \cup J)\ M) = (\int x. (\int y. f\ (\text{merge}\ I\ J(x,y))\ \partial\text{PiM}\ J\ M)\ \partial\text{PiM}\ I\ M)$  (**is**  $?L = ?R$ )  
**and** *integrable* ( $\text{PiM}\ I\ M$ )  $(\lambda x. (\int y. f\ (\text{merge}\ I\ J(x,y))\ \partial\text{PiM}\ J\ M))$  (**is**  $?I$ )  
**and** *AE*  $x$  *in*  $\text{PiM}\ I\ M$ . *integrable* ( $\text{PiM}\ J\ M$ )  $(\lambda y. f\ (\text{merge}\ I\ J(x,y)))$  (**is**  $?T$ )  
 <proof>

**lemma** *product-integral-insert*:

**fixes**  $f :: - \Rightarrow -::\{\text{banach}, \text{second-countable-topology}\}$   
**assumes**  $\bigwedge k. k \in \{i\} \cup J \implies \text{sigma-finite-measure}$  ( $M\ k$ )  
**assumes**  $i \notin J$   
**assumes** *finite*  $J$   
**assumes** *integrable* ( $\text{PiM}\ (\text{insert}\ i\ J)\ M$ )  $f$   
**shows**  $(\int x. f\ x\ \partial\text{PiM}\ (\text{insert}\ i\ J)\ M) = (\int x. (\int y. f\ (y(i := x))\ \partial\text{PiM}\ J\ M)\ \partial M\ i)$  (**is**  $?L = ?R$ )  
 <proof>

**lemma** *product-integral-insert-rev*:

**fixes**  $f :: - \Rightarrow -::\{\text{banach}, \text{second-countable-topology}\}$   
**assumes**  $\bigwedge k. k \in \{i\} \cup J \implies \text{sigma-finite-measure}$  ( $M\ k$ )  
**assumes**  $i \notin J$   
**assumes** *finite*  $J$   
**assumes** *integrable* ( $\text{PiM}\ (\text{insert}\ i\ J)\ M$ )  $f$   
**shows**  $(\int x. f\ x\ \partial\text{PiM}\ (\text{insert}\ i\ J)\ M) = (\int y. (\int x. f\ (y(i := x))\ \partial M\ i)\ \partial\text{PiM}\ J\ M)$  (**is**  $?L = ?R$ )  
 <proof>

**lemma** *merge-empty[simp]*:

$\text{merge}\ \{\}\ I\ (y,x) = \text{restrict}\ x\ I$   
 $\text{merge}\ I\ \{\}\ (y,x) = \text{restrict}\ y\ I$   
 <proof>

**lemma** *merge-cong*:

**assumes**  $\text{restrict } x1 \ I = \text{restrict } x2 \ I$   
**assumes**  $\text{restrict } y1 \ J = \text{restrict } y2 \ J$   
**shows**  $\text{merge } I \ J \ (x1, y1) = \text{merge } I \ J \ (x2, y2)$   
 $\langle \text{proof} \rangle$

**lemma** *restrict-merge*:  
 $\text{restrict } (\text{merge } I \ J \ x) \ K = \text{merge } (I \cap K) \ (J \cap K) \ x$   
 $\langle \text{proof} \rangle$

**lemma** *map-prod-measurable*:  
**assumes**  $f \in M \rightarrow_M M'$   
**assumes**  $g \in N \rightarrow_M N'$   
**shows**  $\text{map-prod } f \ g \in M \otimes_M N \rightarrow_M M' \otimes_M N'$   
 $\langle \text{proof} \rangle$

**lemma** *mc-diarmid-inequality-aux*:  
**fixes**  $f :: (\text{nat} \Rightarrow 'a) \Rightarrow \text{real}$   
**fixes**  $n :: \text{nat}$   
**assumes**  $\bigwedge i. i < n \implies \text{prob-space } (M \ i)$   
**assumes**  $\bigwedge i \ x \ y. i < n \implies \{x, y\} \subseteq \text{space } (PiM \ \{..<n\} \ M) \implies$   
 $(\forall j \in \{..<n\} - \{i\}. x \ j = y \ j) \implies |f \ x - f \ y| \leq c \ i$   
**assumes**  $f\text{-meas}: f \in \text{borel-measurable } (PiM \ \{..<n\} \ M)$  **and**  $\varepsilon\text{-gt-0}: \varepsilon > 0$   
**shows**  $\mathcal{P}(\omega \text{ in } PiM \ \{..<n\} \ M. f \ \omega - (\int \xi. f \ \xi \ \partial PiM \ \{..<n\} \ M) \geq$   
 $\varepsilon) \leq \exp(-2 * \varepsilon^2 / (\sum i < n. (c \ i)^2))$   
**(is ?L ≤ ?R)**  
 $\langle \text{proof} \rangle$

**theorem** *mc-diarmid-inequality-distr*:  
**fixes**  $f :: ('i \Rightarrow 'a) \Rightarrow \text{real}$   
**assumes** *finite*  $I$   
**assumes**  $\bigwedge i. i \in I \implies \text{prob-space } (M \ i)$   
**assumes**  $\bigwedge i \ x \ y. i \in I \implies \{x, y\} \subseteq \text{space } (PiM \ I \ M) \implies (\forall j \in I - \{i\}.$   
 $x \ j = y \ j) \implies |f \ x - f \ y| \leq c \ i$   
**assumes**  $f\text{-meas}: f \in \text{borel-measurable } (PiM \ I \ M)$  **and**  $\varepsilon\text{-gt-0}: \varepsilon > 0$   
**shows**  $\mathcal{P}(\omega \text{ in } PiM \ I \ M. f \ \omega - (\int \xi. f \ \xi \ \partial PiM \ I \ M) \geq \varepsilon) \leq \exp$   
 $(-2 * \varepsilon^2 / (\sum i \in I. (c \ i)^2))$   
**(is ?L ≤ ?R)**  
 $\langle \text{proof} \rangle$

**lemma** (*in prob-space*) *mc-diarmid-inequality-classic*:  
**fixes**  $f :: ('i \Rightarrow 'a) \Rightarrow \text{real}$   
**assumes** *finite*  $I$   
**assumes** *indep-vars*  $N \ X \ I$   
**assumes**  $\bigwedge i \ x \ y. i \in I \implies \{x, y\} \subseteq \text{space } (PiM \ I \ N) \implies (\forall j \in I - \{i\}.$   
 $x \ j = y \ j) \implies |f \ x - f \ y| \leq c \ i$   
**assumes**  $f\text{-meas}: f \in \text{borel-measurable } (PiM \ I \ N)$  **and**  $\varepsilon\text{-gt-0}: \varepsilon > 0$   
**shows**  $\mathcal{P}(\omega \text{ in } M. f \ (\lambda i \in I. X \ i \ \omega) - (\int \xi. f \ (\lambda i \in I. X \ i \ \xi) \ \partial M) \geq$   
 $\varepsilon) \leq \exp(-2 * \varepsilon^2 / (\sum i \in I. (c \ i)^2))$

(is ?L ≤ ?R)  
 ⟨proof⟩

end

## 7 Paley-Zygmund Inequality

This section proves slight improvements of the Paley-Zygmund Inequality [7]. Unfortunately, the improvements are on Wikipedia with no citation.

**theory** *Paley-Zygmund-Inequality*  
 imports *Lp.Lp*  
 begin

context *prob-space*  
 begin

**theorem** *paley-zygmund-inequality-holder:*

assumes *p*:  $1 < (p::real)$   
 assumes *rv*: *random-variable borel Z*  
 assumes *intZp*: *integrable M (λz. |Z z| powr p)*  
 assumes *t*:  $\vartheta \leq 1$   
 assumes *ZAEPos*: *AE z in M. Z z ≥ 0*  
 shows  
 (expectation (λx. |Z x - ϑ \* expectation Z| powr p) powr (1 / (p-1))) \*  
 prob {z ∈ space M. Z z > ϑ \* expectation Z}  
 ≥ ((1-ϑ) powr (p / (p-1)) \* expectation Z powr (p / (p-1)))  
 ⟨proof⟩

**corollary** *paley-zygmund-inequality:*

assumes *rv*: *random-variable borel Z*  
 assumes *intZsq*: *integrable M (λz. (Z z)^2)*  
 assumes *t*:  $\vartheta \leq 1$   
 assumes *Zpos*:  $\bigwedge z. z \in \text{space } M \implies Z z \geq 0$   
 shows  
 (variance Z + (1-ϑ)^2 \* (expectation Z)^2) \*  
 prob {z ∈ space M. Z z > ϑ \* expectation Z}  
 ≥ (1-ϑ)^2 \* (expectation Z)^2  
 ⟨proof⟩

end

end

## References

- [1] G. Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.
- [2] S. Boucheron, G. Lugosi, and O. Bousquet. Concentration inequalities. In O. Bousquet, U. von Luxburg, and G. Rätsch, editors, *Advanced Lectures on Machine Learning, ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany, August 4-16, 2003, Revised Lectures*, volume 3176 of *Lecture Notes in Computer Science*, pages 208–240. Springer, 2003.
- [3] F. P. Cantelli. Sui confini della probabilita. In *Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928*, pages 47–60, 1929.
- [4] B. Efron and C. Stein. The Jackknife Estimate of Variance. *The Annals of Statistics*, 9(3):586 – 596, 1981.
- [5] M. Loève. *Probability Theory I*, chapter Sums of Independent Random Variables, pages 235–279. Springer New York, New York, NY, 1977.
- [6] C. McDiarmid. *Surveys in Combinatorics, 1989: Invited Papers at the Twelfth British Combinatorial Conference*, chapter On the method of bounded differences, pages 148 – 188. London Mathematical Society Lecture Note Series. Cambridge University Press, 1989.
- [7] R. E. Paley and A. Zygmund. A note on analytic functions in the unit circle. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 28, pages 266–272. Cambridge University Press, 1932.
- [8] J. M. Steele. An Efron-Stein Inequality for Nonsymmetric Statistics. *The Annals of Statistics*, 14(2):753 – 758, 1986.